

STIC-Biotech/ChemLib

75771

From: Rao, Manjunath N.  
Sent: Monday, September 16, 2002 9:27 AM  
To: STIC-Biotech/ChemLib  
Subject: Sequence search request for 09/995,587

From: Manjunath N. Rao  
Art Unit 1652, Room 10A11  
Mail Box in Room 10C 01  
Phone: 306-5681

Date: 9-16-02

Please search the following as soon as possible for application with serial number **09/995,587**

**SEQ ID NO: 2 and 10** against all commercial nucleic acid databases including issued patents database and pending application database and provide a **print of all results**.

**SEQ ID NO:1 and 11**, against all commercial protein databases including issued patents database and pending application database and provide a **print of all results**.

If y u have any questions please call me at the above phone number.

Thanks

Manjunath N. Rao, Ph.D.  
Biotechnology Patent Examiner  
Art Unit 1652, Room 10A11  
Crystal Mall 1, USPTO.

RECEIVED  
SEP 16 2002  
STIC/CHEN, Division

No 100% or  
65-100% matches  
for both DNA & protein  
No matches for 15.aa  
fragment was found.

Point of Contact  
P. Sheppard

Searcher: Telephone number: (703) 308-4499  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: \_\_\_\_\_  
Date Completed: 9/27/02  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:

NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_



```

QY 320 FGYNSTAV-----SOEWGSAVLNSDLSIQLFYTRVDSNNTN---HOKIASATL 367
Db 152 FKDNKDLANDPILKNOSEWSSATLSDGQVRLFYTSRQYDPNNLYKGQTLSTAQI 211
QY 368 YLTD-NNGNVSLAQVRNDYIYFE-GDGYYYQTYDQWKATNGADNIARDAHVIEDNGD 425
Db 212 NYSQPDKTLKIDGVEDLKSIDYDGGDKRIYQNVQ--SVGVDNDNHTFRDPHYVED-QGH 268
QY 426 RYLVEASTGLE-NYQGEDQIYNLNGYGGDAFNKSLFRILSNDDIKSRATWANAAIGI 484
Db 269 KYIIFEANTGTETGYGEDSTONPAYYGKMKFFTEBQQNLLQSPK-KGAELANGALGI 327
QY 485 LKLNKDEKNPKVAXLYSPLSAPWSDIERPNVVLGNKYILFAATRLNRGNSDDAWMN 544
Db 328 VELNDY---TLKNVMPPLIASNLVTDETERANVFNKGIWILFTSTRGSKVTVD----- 379
QY 545 ANYAVG-DNVAMVGYVADSLTSGYKPLNDGVLVLTASVPANNRTATYSYAVPVACKDDO 603
Db 380 ---AIGDDDIYMLGYVSTSLTGPYKPLNGTGLVLDHQDLDRDVTWYHAFAIP-OGKGN 435
QY 604 VLVTSYMTNRNGVAGKGM-----DSTWAPSFLIQINPDNTTTLAKWTNQGMWINDSSE 658
Db 436 VVSSYMTNR-----GLFPDHKSTFAPSEFLNKGSKTSYVKNGLILEQGOITVDPTND 488

RESULT 2
Q97181
ID Q97181 PRELIMINARY; PRT; 489 AA.
AC Q97181
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DE LEVANSUCRASE.
GN CAC1172.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=2139325; PubMed=11466286;
RA Noelling J., Breton G., Onelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe P., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT *Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.*;
DR J. Bacteriol. 183:4823-4838(2001).
DR EMBL: AE007686; AAK79737.1;
DR InterPro: IPR003469; Glyco_hydro_68.
DR Pfam: PF02435; Glyco_hydro_68; 1.
KW Complete proteome.
SQ SEQUENCE 489 AA; 54553 MW; 9002CB364F1D3CD8 CRC64;

```

```

Query Match 17.1%; Score 707; DB 16; Length 489;
Best Local Similarity 37.2%; Pred. No. 9.5e-25;
Matches 183; Conservative 88; Mismatches 173; Indels 48; Gaps 20;
QY 188 LTDDQIKALKNKFSKAASKTQMTYNDFOKRIADTLIKQDGRYVFPFKASEIKNMPAAT 247
Db 21 ITQQTFASTNDMT-KETGYVSHITRYNMSKI--PHEQNDLKFKVPQFNASTLKNASAK 77
QY 248 TKDAQWTNIEPLDWDSPVQDVRTGOVANNQYQLVIAAMGIP-NQNDNHLYLLNKGK 306
Db 78 GYDKNGNLID-LDWDSPNPLON-GDGTVANYGHYVIFALAGDPKNQDTSIMTYQKIT 135
QY 307 DNELSHWKNVGPFGYNSTAV-----SOEWGSAVLNSDLSIQLFYTRVD--TSDN 355
Db 136 ENSIDSWKNAGVKFSDSKRYVANDPYLKYQTEWSSGATLSGQVRLFYTFDGSVAKDG 195
QY 356 NT--NHQKIASATLYLTDNNGN-VSLAQVRNDYIYFE-GDGYYYQTYDQWKATNK--GAD 409

```

```

Db 196 GTDASNQVITTTQVNLSDPSNTINISVSDHKSVPDGGNGTIYQNVQOIFDEGKWSSGD 255
QY 410 NIAMRDAHVIEDONGRDRYLVEASTGL-ENYQGEDQIYNLNGYGGDAFNKSLFRILSN 468
Db 256 NHTLRDPHYVED-NGRKYLVEANTGTNDYOGDTSLLNKAFYGRSQSFETKQDLID 314
QY 469 DTKSRATWANAAIGILKLNKDEKNPKVAXLYSPLSAPWSDIERPNVVLGNKYILF 528
Db 315 TNKKHDASLANGALGIENNDYTLKKEMK---PLIASNTVTDIERANVFNKMGWYLF 371
QY 529 AATRLNRGNSDDAWMNANYAVGDNVAMVGYVADSLTSGYKPLNDGVLVLTASVPANNRTA 588
Db 372 TD---SRGSK-----MTINGISSKDIYMLGPFSSNLSTGPKPLNKTGLVNLNLDPTDLTF 424
QY 589 TYSYAVP-VAGKDDOVLVTSYMTNRNGVAGKGM-----DSTWAPSFLIQINPDNTTTLA 643
Db 425 TSHFAPVPTNGKN--VVITSYITNR-----GMYSDDHSSFAPSFLNKGSKTSVSN 476
QY 644 KMTNOGDWTD 655
Db 477 SILQGGQLTIDN 488

RESULT 3
Q82854
ID Q82854 PRELIMINARY; PRT; 487 AA.
AC Q82854
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DE BETA-FRUCTOFURANOSIDASE.
GN CAC1172.
OS Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1409;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V230;
RA Tsuchiaki K.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=V230;
RA Kurimoto M.;
RT "Cloning and sequencing of beta-fructofuranosidase gene from Bacillus
RT sp. V230.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB010272; BAA32083.1;
DR InterPro: IPR003469; Glyco_hydro_68.
DR Pfam: PF02435; Glyco_hydro_68; 1.
SQ SEQUENCE 487 AA; 53412 MW; 7F52A47921824AE3 CRC64;

```

```

Query Match 14.9%; Score 613.5; DB 2; Length 487;
Best Local Similarity 34.0%; Pred. No. 1.6e-20;
Matches 168; Conservative 81; Mismatches 158; Indels 87; Gaps 21;
QY 199 MNFSAKASGT-----QMTYNDPQK-----IADTLIKQDQ-- 228
Db 1 MNFKRLAKAAAVTFTAILVAGDGPPIFAQOMNSGDKYDGFHAHTRAD-MLKIPQQQ 59
QY 229 ---RYTVPFFKASEIKNMPAATTTDAQNTTIEPLDWDSPVQDVRTGOVANNQYQLV 285
Db 60 NSPOKVPQFNASAIKNIDSAGYDKSGNLID-LDWDSPNPLON-DGTAANTYGHYHVS 117
QY 286 AMMGIP-NQNDNHLYLLNKGYNELSHWKNVGPFGYNSTAV-----SOEWGSA 335
Db 118 ALAGDPKNSDDTFLHLYFYQVGTSDSWKNAGRVFEDMKFVPNDPYLKYQTEWSSGA 177
QY 336 VLNSDLSIQLFYTRVDTSDN-----NTNHQKIASATLYLTD-NNGNVSLAQVRNDYIYF 388
Db 178 TLTKDQVRLFYT--DYSGNPDGGTGAGNQIISTAQVNLSDPSNTINISVSDHKSVP 235

```

QY	389	E-	CDGYYTYDQWKATNK--GAONIAMRD	AHVIEDGNGDRYLVFEASTG-LENYOGEDQ	444
Db	236	DGGDGT	YQNTQOFTDEGKWISGONHTLRPHYVE--KGHKYLVFEANTGTTDYGQDQS	294	
QY	445	IYNWLN	YGDDAFNTKSLFRILSND	DIKSRATWANAAGILKLNKDEKPNKVAEL	SPLI 504
Db	295	FNKKAY	YGSDVFFONERKNILQSPK-KQIASLANGALGIVELADDY	---TVKSYMKPLV	350
QY	505	SAPVMS	DIERNVYKLGNYLFAAATRLN	RGNSNDAMNANYAVGD-NVAMVGYVADSL	563
Db	351	ASNTVAD	EVERANIFKMNKLYFLTDSRGSKMTSD-----GINDKDYWMLGPGDSSL	402	
QY	564	TGSYKPL	NDSGVLTASVPANRWTATSYAVP	VAGKDQVLTYSYMTNRN-----	614
Db	403	NGPHNP	INETGLVNMNLDPADLTHYSHCGIP-HPEGNNVYLTSYMTNRGYP	PEHSHL	461
QY	615	---GV	AKGMDST	624	
Db	462	RDKLGV	NIKGS	DTs 475	

RESULT	4
Q97I79	PRELIMINARY; PRT; 428 AA.
ID	Q97I79
AC	ATC
DT	01-OCT-2001 (TReMBLrel. 18, Created)
DT	01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT	01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE	LEVANSUCRASE.
GN	CACL1774.
QS	Clostridium acetobutylicum.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC	Clostridium.
QC	NGBL_TaxID=1488;
QN	[1]
RP	SEQUENCE FROM N.A.
RX	STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
RC	MEDLINE=21359325; PubMed=11466286;
RA	Noelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.,
RA	Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA	Tatsov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA	Bennett G.N., Koehn E.V., Smith D.R.;
RT	"Genome sequence and comparative analysis of the solvent-producing
RT	bacterium Clostridium acetobutylicum".
RL	J. Bacteriol. 183:4823-4838(2001).
EMBL:	AEO07686; AAK79739.1; -
DR	InterPro; IPR003469; Glyco_hydro_68.
DR	pfam; PF02435; Glyco_hydro_68; 1.
KW	Complete proteome.
SC	SEQUENCE 428 AA; 48395 MW; 9670B154B178E23E CRC64;

Query Match	8.6%	Score	356.5;	DB	16;	Length	428;
Best Local Similarity	27.2%;	.Pred. NO.	5.4e-09;				
Matches	125;	Conservative	70;	Mismatches	176;	Indels	89;
						Gaps	22;

[illegible]

	QY	425 DR-YLVFEASTG--LENYOGE---DQIY--NWLNYGGDDAFNIKSLFRILLSNDODIKSRAT	476
	Dd	232 KKYLIIEFEGNKGGKIEKMKNENIGDKLFRKHHTAPRGVENFN-----	273
	QY	477 WANAAGILINKLNDEKPNAEYLSPPLISAPMVSDETERPNVVKLGNNKKYYLFATRLNRG	536
	Dd	274 -GNVGIAVAQ-NDKLTRFK---LLPPLERAVGVNQQLRQPQIVMKNNKKYYLTISH----	324
	QY	537 SNDDAMWNAVAYVDN--VAMVGYSADLSFGSKPKLDNSGWILTASVPANWRRTATSYYA	594
	Dd	325 -----KETYAOLGANVGDLGYFCGCSGLRSNTKPLNGNLVLTN--PTNDPQTITSWYL	375
	QY	595 VPVAGKDQQVLVTSYMTRNRNVAGKGMDSTWAPSFLLOIN	634
	Dd	376 --VSGHDVLSFINEY--HFNQGRLRYG--GTFAFT--LQIS	407
	RESULT	5	
	Q9EVD6	PRELIMINARY;	PRT;     630 AA.
	ID	Q9EVD6;	
	AC	Q9EVD6;	
	DT	01-MAR-2001 (TREMBLrel. 16, Created)	
	OT	01-MAR-2001 (TREMBlrel. 16, Last sequence update)	
	DT	01-DEC-2001 (TREMBlrel. 19, Last annotation update)	
	DE	FRUCTOSYLTRANSFERASE.	
	DN	Ftp:	
	GS	Actinomycetes naeslundii.	
	QC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;	
	OX	Actinomycetales; Actinomycetaceae; Actinomycetaceae; Actinomycetes.	
	CC	NCBJ_TaxID=1655;	
	RN	[1]	
	RP	SEQUENCE FROM N.A.	
	RC	STRAIN=ATCC 12104, WVU45;	
	RA	Bergeon L.J.; Moroi-Bermudez E., Burne R.A.;	
	ET	"Characterization of the fructosyltransferase gene of Actinomycetes naeslundii."	
	RL	Submitted (JMBL-2000) to the EMBL/GenBank/DDBJ databases.	
	DR	EMBL: AF228582; AAG09737.1; "	
	DR	InterPro: IPR003469; Glyco_hydro_68.	
	DR	Pfam: PF02435; Glyco_hydro_68; 1.	
	KW	Transferase.	
	SQ	SEQUENCE	630 AA; 68256 MW; CCA0A04184DD6D7D2 CRC64;

[illegible]



```

Db 313 IVQSEGRYATKN---GVVLGFTQHOLLVDPGKIYQTRQ-----NPG-----VNFROPPTF 363
Qy 420 EDNG-----DRIYLFEEA-----STGLNYQGEDOIYNWLNWYGGD-- 454
Db 364 RDQNNPSDPTPEYVMFEGNSAFVREQQVYVDAARAGQNTTLATCEEDLGYE-----KGDPK 419
Qy 455 ----DAFNKSLFRILSNDDIKSRATWANAAIGILKLNDEKNPKVA---ELYSPLISAPM 508
Db 420 AETVEAVNORGGYYQLAN-----VGLAR-----AKNKAMTEWEYLPPLLSGNC 462
Qy 509 VSDEIERPNVVKNGKYYLFAAPRLNRGSDNDAMNANTAVG--DNVAMVGVVADSLTGS 566
Db 463 VNDOTERPOIYQDQGYLYFTISH-----RETADGLOGPEGYGFVGGGLRSD 511
Qy 567 YKPLN-DSGVVLVTASVPANRRTA-----TYSYAVPVACKDDQVLVTYSY 610
Db 512 YKPLNQNTGALGNLNFNFGKPYSPDNQSPYTFQSYSHYVMP--GGLVESFIDSIG 569
Qy 611 TNRNGVAGKGMDSWAPSLLOINPDNTTTLAKMTN 647
Db 570 GNKGDPNVRG--GSLSPYVKLNISGDTSDVDRYGTN 604

RESULT 6
Q93FU9 ID Q93FU9 PRELIMINARY; PRT; 424 AA.
AC Q93FU9;
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE LEVANSUCRASE.
OS LSCA.
OC Pseudomonas aurantiaca.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=86192;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S-4380;
RA Koh I., Jang E.K., Kim J.S., Song K.S., Kim C.H., Rhee S.K.;
RT Screening, molecular cloning and characterization of a novel
RL levanucrase, and its expression in Escherichia coli.;
RL Submitted (Sep-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF306513; AAL09386.1;
SQ SEQUENCE 424 AA; 47009 MW; 9C1B4936A985162E CRC64;

```

```

Query Match 6.8%; Score 279.5; DB 2; Length 424;
Best Local Similarity 24.3%; Pred. No. 1.6e-05;
Matches 118; Conservative 76; Mismatches 174; Indels 113; Gaps 25;

Qy 197 NKNMFSKAAKSGTOMTYNDPQKI-AOTLIKQDGRYTVPFKASEIKNMPAANTKDAQWT 255
Db 4 NTEKFGKAPHQPSLWTRADALKVHAD-----DPTTQPLVSA-----DFPVLSE----- 48

Qy 256 IEPLDVWDSPPQDVRTGOVANNWNGVQLVIAMG--IPNQ-----NDNH 297
Db 49 ---VFIDWTHPLKDL-DGHNTSYDGSVFTLTADRHNDPEYIDENGNYDITRDNDH 104

Qy 298 ----IYLYNKYGDNLHSHWKNVGPVFGYNSTAVSQWGSVAV-LNSDNIOLFTRVDT 352
Db 105 GRAKMYWFESRTGK----WKLGRVMAEGVSPTRAWAGTPLLNEGDVDLYYTA-- 158

Qy 353 SDNTHNHQIASATLYLTNNNGVSLAQVRNDYIVFEGDGYIYQTDQ---WKATNKAGD 409
Db 159 ----TPGATIVKVRGRVVTTEHGVSVMGFEKVKPLFADGKMYQTEAQNPFW----- 206

Qy 410 NIAMRDARHVEDGN-GDRVLVFEASTGLNTEGQEDOIYNWLNWYGGDDAFNKSFRILSN 468
Db 207 --GFRDPWFDPNDGKLYHMFEGNVAGE--RGSHK-----GKAETGDVPPGYEDVGN 256

Qy 469 DDIKSRATWANAAIGILKLNDEKNPKVAEYLSPLISAPMVSDEIRPNVVKNGKYYLF 528

```

```

Db 257 -----SRFQACVGI-AVARDEGDD-WEMPLPLLTAVGVNDQTERPHFVQDCKYYLF 308
Qy 529 AATRLNRGSDNDAMNANTAVGDNV-----AMVGYVADSLTSGYKPLNDGSGVVLVTASVPAN 584
Db 309 T-----ISHTFTYADGVTGPDGVYGFVADSLFPGPYPLNGSGLVL--GNPSS 353
Qy 585 WRTATYSYAVPVACKDDQVLVTYSY-----TNRNGVAGKGMDSWAPSLLOINPDNTT 640
Db 354 QPFOFYSHCVMP-----NGLVTSFIDSVPTDESQTQIR-IGGTEATVEIKIKGQQTFFV 406
Qy 641 V 641
Db 407 V 407

RESULT 7
Q9AEX9 ID Q9AEX9 PRELIMINARY; PRT; 415 AA.
AC Q9AEX9;
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DE 01-OCT-2001 (TremBLrel. 18, Last annotation update)
DE LEVANSUCRASE.
OS Rannella aquatilis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Rahnella.
OX NCBI_TaxID=34038;
RN [1]
RP SEQUENCE FROM N.A.
RA Kim H., Cha J.;
RT "Cloning and Characterization of levanucrase.";
RL Submitted (Feb-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY027657; AAK14794.1;
DR InterPro; IPR003469; Glyco_hydro_68.
DR Pfam; PF02435; Glyco_hydro_68; 1.
SQ SEQUENCE 415 AA; 45938 MW; F09F5A1BF8690C5E CRC64;

```

```

Query Match 6.6%; Score 271.5; DB 2; Length 415;
Best Local Similarity 26.6%; Pred. No. 3.6e-05;
Matches 102; Conservative 60; Mismatches 132; Indels 89; Gaps 19;

Qy 261 VMDSWPQDVRTGOVANNWNGVQLVIAMGIPNQNDNHYL-----LYNKYGDNE 309
Db 44 INDAMPLRSL-DGTVVSDGNSVFTLTQARN-NNNSEYLDADGNYDITSDNNNRHGRK 101

Qy 310 LSHWKN-----VGPFGYNSTAVSQWGSVAV-LNSDNIOLFTRVDTSDNTHNQ 360
Db 102 ICYWSRTGKDWIFGGRVMAEGVSPTRWAGTPLLNEGDVDLYYTCV-----TPGA 155

Qy 361 KIATATLYLTNDNNGVSLAQVRNDYIVFEGDGYIYQTDQ---WKATNKAGDIAMRD-A 416
Db 156 TIAKYGRKVLSEECVTLAGFNEVKSLSFADGVYIYQTESQNTYWN-----FRDPS 205

Qy 417 HVIDGNGDRVLVFEASTGLNTEGQEDOIYNWLNWYGGDDAFNKSFRILSNDDIKSRAT 476
Db 206 PFIDPHDGLKLYMVEGNAVGE--RGSHVI-----GKQEMGTLPDHRDVG-- 251

Qy 477 WANAAGITL---KLNKDEKNPKVAEYLSPLISAPMVSDEIRPNVVKNGKYYLFAATRL 533
Db 252 YQACGIGNAVAKDLSGDB-----WEILPPLVTAVGVNDQTERPHFVQDCKYYLFTISH- 305

Qy 534 NRGSDNDAMNANTAVG-----DNVAMVGVVADSLTSGYKPLNDGSGVVLVTASVPANRRTAT 589
Db 306 -----KFTYAEGLTGPDGV--YGFISDRLTGYPSPWNGSGLVL--GNPSPQFPQT 351

Qy 590 YSYAVPVACKDDQVLVTYSMTN 612
Db 352 YSHCVMP-----NGLVTSFIDN 368

```

RESULT 8

[illegible]

```

QY 430 FEASTGLENYQGGDIYNLWNGDDAFNFKSLFRILSNDDIKSRATWANAAGILKLNK 489
Db 467 -----TELYGYP-ERAYKSYVYGG-----YRLTWNGNL--VLYSNK 500
QY 490 DEKNPKVAEL-----YSLISAPMVSEIERPNVVKLGNYLYFAATRLNRGS----- 537
Db 501 ADGNGKNGIIOQNDPEYKEDTAKGTMSGOYDAKQIE-}-----TEEQDNTPLDI 550
QY 538 -----NDADWNNANY--AVGDNVAMVG----- 557
Db 551 DIHTAIDGEGYVDGYIETIEDTSSAIDIDYHTAVDSEVHGVCYTESSEBSNPIDEEE 610
QY 558 -----XVADSLTGSYKPLNDSGV-----LTASVPANWRTATYSY 593
Db 611 STHENSKHHADVVEEEDNPGGQVTTESNLVEFDEESTKGIVTGAVSDHTTIEDTKEY 670
QY 594 AV-----PVAGKDDQVLVTSYMTNRNGVAGKGMDSWAPSLQLINPDN 637
Db 671 TTESNLIELVDELPEHQAQGPBIEITENNHHISHSLGTENGHGNV--GVIEEIEBNS 728
QY 638 TTVTLAKM-----TNQGDWIDDSSENLDN-----IGLDSAALP----- 672
Db 729 HVDIKSELGEGSGNQSGNQSFEETEE--DKPKYEGGQGNIVDIDFDSVPQIHQNGQDSF 787
QY 673 ---GERDKP-----VMDLIG--YGLKPHD----- 692
Db 788 EDTEDKPKYEHGNGNIIDIDFDSVPQIHGFKNHIEEEDTNKDKPNYQFGGHSVDPE 847
QY 693 -----PATPNDET--PTTPTPETPNTPKTPKTPEN 722
Db 848 EDTLPKVSQNEGQQTIEDTPTPTPTPEVPSEPTPTPTPEVPSEPTPTPTPEVP 906
QY 723 PGTQPTPNTPPEIPL-----TPETPKQPETQT----- 751
Db 907 PSEPTP--TPPTPEVPSEPTPTPTPEVPSEPTPTPTPEVPAEPGKPVPAKEPKK 965
QY 752 -----NNRLPQTG--NNANKAMTGLCMGTLLS 776
Db 966 PSKPEQGVKVTPIETNEKVAAPTAKAQSKESELPETGGEESTNKGML---FGGLFS 1022
QY 777 MGLAEINKRRFN 789
Db 1023 ILGLALLRRNKN 1035

RESULT 10
Q53682
ID Q53682 PRELIMINARY; PRT; 940 AA.
AC Q53682;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE FIBRONECTIN BINDING PROTEIN B.
GN FNBB
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=8325-4;
RX MEDLINE=92111475; PubMed=1837266;
RA Jonsson K., Signas C., Muller H.P., Lindberg M.;
RT "Two different genes encode fibronectin binding proteins in
RT Staphylococcus aureus. The complete nucleotide sequence and
RT characterization of the second gene.";
RL Eur. J. Biochem. 202:1041-1048(1991).
DR EMBL; X62992; CAA44726.1; -
DR InterPro; IPR004237; Fn_bind.
DR Pfam; PF02986; Fn_bind; 2.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.

```

SQ SEQUENCE 940 AA; 103555 MW; E35FBBCA907AE345 CRC64;

Query Match 6.0%; Score 246; DB 2; Length 940;  
Best Local Similarity 18.6%; Pred. No. 0.0013;  
Matches 185; Conservative 135; Mismatches 334; Indels 340; Gaps 44;

```

QY 55 VTLLGSGQ-----VSAASDTTITSANANSASAAANTQNSOVASSAATSTSSAASLN 109
Db 25 IYVGQGEKEAAASQNNNTVEESGSSATESKASETQTTNNVNT---IDETQSYSATST 81
QY 110 NTDSKAAQENTNTAKNDTOKAAPANESSEAKNEPAVNVND--SSAANKNDQOQS----- 162
Db 82 EQPSQSTQVTTTEAPK---TVQAPKVTESRV-DLPSEKVADEKTTGTQVDTIAQPSNVSEI 137
QY 163 ---KKNT--TAKLNKDAENVVKAGIDPNSLTD-----DQIK 194
Db 138 KPRMKRSTDTVTAAVEKEVEETKATGTDVTKNVEEGSEIVGHKQDVTNVNPHNAERTV 197
QY 195 ALNKMNFSSKAASGTQMTYNDQKIADTLIKQDGRYTVPEFFKASEIKNPAATTKDAQTN 254
Db 198 LKYKWKCEGKAG--DYEDF-TLSD-----NVETHGISLTKRVPKIKSTGQV- 243
QY 255 TIEPLDWDSPVQDVRTGQVANNN-----GYOLVIAMGIPNQN 294
Db 244 -----MATGEIIGERKVRYTEKVEYVQEKKDLTAELSLNLFIDPTVTQKG 288
QY 295 DNHIYLLYKNYGDNELSHWKNVGPFGYNSTAVSQEKSGLAVLNSDNSIQLFYTRVDT-- 352
Db 289 NONVEV---KLGEITVSKIFNIQVLLG---VRDNGVTA-----NGRIDTLN 329
QY 353 -SDNTHNHQKTASATLYLTDNNGNVSLAQVNDYIVFEGDGYIYQTDQMKATNKGADNI 411
Db 330 KVDGKFH-----FAYMKPNQSLSSVTGTQVTKGKPKVNNPTVKYK--HIGSDDL 381
QY 412 A-----MRDAHVIEDGNGDRLVFEASTGLE-NYQGEDQIYNWL-----NYGGDDAFNIK 460
Db 382 AESVYAKLDDVSKFEDVDNKSLOFDITNGGYSLNFNLDQSKNYVIKEYGIDSNASNL 441
QY 461 SLFRILS--NDDIKSRATWANAAGILKLNKDEKNPKVAELYSPLI--SAPMVSEIERP 516
Db 442 FQHLFGYNYIYNTSLTNKWN--GVAFYSNNAQGDGKDLKEPIIEHSPIELFEFSEP 498
QY 517 NVKLGKNKYLLFAATRLNRGSDNDAWNNANY-----AVG----- 550
Db 499 PVKE-----HELTGTIEESNDSKPIDFEYHTAVEGAEGHAETIETEDSIHVDFEES 551
QY 551 ---DNVAMGVYADSLTGSYKPLNDSGV-----LTASVPANWRTATYSYA 594
Db 552 THENSKHHADVVEEEDNPGGQVTTESNLVEFDEESTKGIVTGAVSDHTTIEDTKEYT 611
QY 595 V-----PVAGKDDQVLVTSYMTNRNGVAGKGMDSWAPSLQLINPDNT 638
Db 612 TESNLIELVDELPEHQAQGPBIEITENNHHISHSLGTENGHGNV--GVIEEIEBNSH 669
QY 639 TTVLAKM-----TNQGDWIDDSSENLDN-----IGLDSAALP----- 672
Db 670 VDIKSELGEGSGNQSGNQSFEETEE--DKPKYEGGQGNIVDIDFDSVPQIHQNGNQSPF 728
QY 673 ---GERDKP-----VMDLIG--YGLKPHD----- 691
Db 729 EDTEDKPKYEGGNGNIIDIDFDSVPQIHGFKNHIEEEDTNKDKPNYQFGGHSNVDPEE 788
QY 692 -----DPATPNDETPTTPTPETPNTPKTPKTPENPGTPTPTPTPTPTPT 733
Db 789 DTLPQVSGHNEGQQTIEDTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 846
QY 734 TPEIPLTP-----ETPKQPE----- 757
Db 847 TPEVPTPEPKPIPPAKEEPKPSKPEQGVKVTPIETNEKVAAPTAKAQSKESELP 906
QY 758 TG--NNANKAMTGLCMGTLLSMFGLAEINKRRFN 789

```

QY	179	KK-----AGID----	PNSLTDDOI	KALNKNFSAKSGTQM	TYNDF-----OKIAD--	221
Dd	270	KKETSSBNTQKVDEHYANS	LQLAQSLEEDKATTNEQA	TVK-NOFLNAQKLKEIQ	328	
QY	222	TLIKDGRYTVPPFKA-----	SEINKMPAAATKDAQTNT	TEPLDVMS-----	264	
Dd	329	PLIKET---NVKLYKAMBS	LEQVEKELKHNSEANLED	VAKSKEYREYEGKLNOSKNL	385	
QY	265	-----	WPVODYRTGQVANWNGYQ	--LVIAWMGI	290	
Dd	386	PELKQLEEBAHSLKOVVED	FRKKFTSEQVTPKRKYRDL	AANNENQOIKIELTVSPENI	445	
QY	291	PNQNDNHIIYLNYKGDN	LSHWNVGPIFGVNSTAVSQEWS	GSAVLNSDNSQLPTTRV	350	
Dd	446	TVTEGEDYKFTVIKSDSKTT-	LDPSDLLTKYPSPVSDRI	SNTKYKTNTDN-----HKIA	498	
QY	351	DTSDONTNQKTSATATLI	YTDNNHNVSUAQVRNDYI	VFEGDGYVYTQDKWATNKADN	410	
Dd	499	EITIKMLKINESQTVTLA	KAKDSGVN----VEKTFITV	-----QKKEEKVPKPT	544	
QY	411	IAMDRAHVIE-----	DGNDRYLVFEASTGLENY	QCGEDIYNWLNVGGDDAFNIKS	461	
Dd	545	PEQKDSKTEKYPQEPKSN	DKNQNOLIKLSAQELEKLE	-----KA	585	
QY	462	LPRILSNDIKS-----	RATWAN-----	AAIGTLKLNKDEKNPKVAELYSP	507	
Dd	586	IKELMEOPEIPSPNPEYG	IOKSIWESQKEPIQEAIT	SPKKIIGDSSSKYYTEHYFNKYKD	645	
QY	508	MYSDEITERP-----	NVVKLGANKYYLFAATPL	NRGNSDDAMNANTAYAGNVAMGYVAD	561	
Dd	646	FNNYQLHAQMEMLFRKY	QQVYNNKPPDNAEIKKIF	ESDMRKTBKBDNYGSLENDALKGYFEK	705	
QY	562	SUTGSYKPL-----	NDSGVYLTFASYPANWR	TATSYZAVPVAGKDDQ-----VLVTSYMT	611	
Dd	706	YELTFPNKIKQIVDDLO	KKVBQDOPA-----	PIPENSEMDQAKEKAIASKVMS	755	
QY	612	NNRGVAGMGDMSTWAP	SFLLIQINPDNTTYTLAK	NTNGDWINDSSSENIMTGDLDSAAL	671	
Dd	756	-----KVLGVHQ-----	HLOKKNHISKIV-----	DLFKELEAIKQOTTFIDDNAKT	796	
QY	672	PEERKPKVDMDLIG-----	YGLK-----PHDPATPN	DPDTPPTPETPETPNTPKTPK	718	
Dd	797	EVEIDNLVH-DAFSKMN	ATVAKFOKGLENTNPET	PTDTPKIPELPQADPTPQADPTPHVPE	855	
QY	719	TPENPGTPTPTNTPTPE	ILPTETPKQPCTQTNRLPOT	758		
Dd	856	SPKAEPAPRPESPCKT	PEAPHVPESPCKTPEAKPI	EPKPPT 895		
RESULT	12					
Q9LCJ7						
ID	Q9LCJ7	PRELIMINARY;		PRT; 1016 AA.		
AC	O9FCJ7					
DT	01-OCT-2000 (TrEMBLrel. 15, Created)					
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)					
DT	01-DEC-2000 (TrEMBLrel. 19, Last annotation update)					
DE	DETRANSUCRASE.					
GN	DSRT.					
OC	Leucostoc mesenteroides					
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;					
OX	Leucostoc.					
NCBI_Taxid=1245;	[1]					
SEQUENCE FROM N.A.						
RC	STRAIN=NRRL B-512F.					
RA	MEDLINE=20169623; Pubmed=10705445;					
RA	Funane K., Mizuno K., Takahara H., Kobayashi M.;					
RT	"gene encoding a dextran sucrase-like protein in Leuco					
RT	mesenteroides NRRL B-512F."					
RL	Biosci. Biotechnol. Biochem. 64:29-38(2000).					
DR	EMBL; AB020020; BAA90527.1; -					
DR	HSSP; P06278; 1VJS.					
DR	InterPro; IPR003318; Glyco_hydro_70.					

```
DR Pfam; PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1016 AA; 110344 MW; 8896EFDE13CCB47 CRC64;

Query Match          5.7%; Score 236.5; DB 2; Length 1016;
Best Local Similarity 21.2%; Pred. No. 0.0037;
Matches 162; Conservative 103; Mismatches 266; Indels 233; Gaps 37;

QY 1 MYKSGKNAVYTLSTAALVGCATVNASADNTENN-----DSSTVQVTVGNDIA 51
DB 1 MYKSGK-----MLVIAGSVIIIGVTFIOQAQADVQNGVAVATVPQSNLDATSD- 53
QY 52 VKSVTLASGQVQASAD-----TIRTSANANASSAANTONSNOVASSAATSTSSA 105
DB 54 -KSITDDKAAATATSTDDKATTTADTSTDDKAAATAA--STDDKATTTATSTDDKAT 110
QY 106 ASLNTDSKAA-----QENTNTAKNDTQKA-----APANESSEAKNEPAYNVNDSSA- 153
DB 111 TAATSTDDKAAATATSTDDKATTTAATSTDDKATTTAATSTDDKATTTAATSTDDKAA 170
QY 154 ----AKNDQ-----OSSKNTTAKLNDKAENNVKKA-GIDPNSLTDDQIKALNKWFSK 203
DB 171 TADTSTDDKATTTAGTSDNNNTASDKVSSSAQKSQSIDNNSKTADTTAALEASSKNL 230
QY 204 AAKSGTQMTYNDQFKIADTLIKODGRYTPFFKASEIKNHPAATTKDAQNTNIEPLDVM 263
DB 231 KTDGKTYIDD-----DQVK-----KNF--ATVIDGKVLVYF----- 261
QY 264 SWPVQDVRGQVANWNGYQLVIAMGIPNQNDNHIYLLNKGDNELSHWKNVGPFGYN 323
DB 262 ----DKETGALADTNDYQFLEGL-----TSENWYTHEANVSQVTSYTNVD--GY- 307
QY 324 STAVS-----QEHSGSAVLNDSNSIQLFYTRVDTSDNNTNHQIKASATLYLTDN 372
DB 308 LTADSWYRPFKILVNGQNWESSKDDLRPLMTWPKATQVNYLNAMKYLDAETET-- 365
QY 373 NGNVSLAQVRNDYIVFEGDGYIYQYDQWKATNKGADNIARDHVIEDGNGDRL- 428
DB 366 ----VFEASTGLENYQSGDQIYNWLNLYGGDQAFNKLFLRLSNDQIKSRATW 477
QY 429 ----VFEASTGLENYQSGDQIYNWLNLYGGDQAFNKLFLRLSNDQIKSRATW 477
DB 405 SKFVDSQSNWNIASEKSTDLHQGALLY-----VNSDKTPDANSYRL----- 448
QY 478 ANAATGILKLNKDEKNPRVAEYLSPLISAP-----MVSELERPNVV-----KLGKYY 526
DB 449 ----LNRTPNQGTGLPT--TDPTQGYDFELANDVDSNPVYQAEQLNWMY 496
QY 527 LFAATRLNRGS--NDDAMNANY-----AVGDNV--AMGYVADSLTGSYKPLNDSGVV 576
DB 497 L-----LNFSGITNDA--DANFDSIRVDV--DNVDADLLQIAADYFKAAY-GVDKSDAI 547
QY 577 LT--ASVPANWRATSYAYAVPVAGKDDQVL-----VTSYWTRNGV 616
DB 548 SNQHVSTLEDWSDNDAY-----VKDNGDNLSDMNKRLSLKYSLPMFPAVDQYGNKRSGL 603
QY 617 ----ACKGMDSTWAPSF-LIQINPDNTTTLAKMTNQ 648
DB 604 EPFTNSLVDRDNTDNTAQPNSYSEVRAHDSQVQVIAELIKQ 647

RESULT 13
Q99RD3 ID Q99RD3 PRELIMINARY; PRT; 961 AA.
AC Q99RD3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE FNBB PROTEIN (FIBRONECTIN-BINDING PROTEIN HOMOLOGY).
GN FNBB OR SA2290 OR SA2502.
OS Staphylococcus aureus (strain N315), and
GS Staphylococcus aureus (strain Mu50).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
```

OC Bacillus/Staphylococcus group; Staphylococcus.  
OX NCBI\_TaxID=158879, 158878;  
RN [1]

RP SEQUENCE FROM N.A.  
RC SPECIES=S.aureus (strain N315), and S.aureus (strain Mu50);  
RX MEDLINE=21311952, PubMed=11418146;

RA Kuroda M., Ohca T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,  
Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
Sekimizu K., Hirakawa H., Kubara S., Goto S., Iabuzaki J.,  
Kanthisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;  
RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
aureus";

RL Lancet 357:1225-1240(2001).  
DR EMBL; AP003137; BAB43593.1; -;  
DR EMBL; AP003137; BAB43593.1; -;

DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR002965; P-rich\_extensn.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR PRINTS; PR01217; PRICHEXTENS.

DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
DR Complete proteome.  
KW SEQUENCE 961 AA; 106010 MW; 364940F884EA4101 CRC64;

Query Match 5.6%; Score 231.5; DB 16; Length 961;  
Best Local Similarity 19.9%; Pred. No. 0.0059;  
Matches 199; Conservative 121; Mismatches 349; Indels 331; Gaps 52;

QY 55 VTGSGQ-----VSAASDITRTSANANASSAANTONSNOVASSAATSTSSAASLN 109  
DB 25 IVVGMGQKEAAASEQNNTTVEESGSSATESKASEQTNTNNVT---IDETQSATST 81

QY 110 NTDSKAQENTNTAKNDTQKAPANESSEAKNE--PAVNVD-----SSAA 154  
DB 82 EQPSKSTQVTTTEA---PTTVQAPKVETEMKSOEDLPSEKADVADKETTQVQDIAQPSNV 138

QY 155 KNDQSSKNTTAKLNDKAENNVKAGIDPNS---LTDQIKALN----- 197  
DB 139 EIKPRMKRSADVTAVEKEVAEEAKATGDTVTKNKEVTESSLEGHNKDSNVNPHNAQVR 198

QY 198 ----KNFSAKSGQMTYNDQFKIADTLIKODGRYT---VPFFKAS-EIKNM----- 243  
DB 199 TLKYKKKFGEGKAG---DYDFD-TLSDN-VETHGISTLRKVPKESKTEDEKVMANGQVI 253

QY 244 ----PAATTKD-----AQTNTIEPLDWDSD 264  
DB 254 NERTIRYTFDYINNKDLTAELNLNLFIDPTVTIKQSQKVEVTIGQNKYSKEFDIKYL 313

QY 265 WPVQDVRGQVANWNGYQLVIAMGIPNQNDNHIYLLNKGDNELSHWKNVGPFI-FCYN 323  
DB 314 DGVKD-RMGVTVNGR-----IDTLNKEEGKFSHFAYVKNQSLTSVTVTQVTSYK 365

QY 324 STAVSQWSSGSAVLNDSNSIQLFYTRVDTSDNNTNHQIKASATLYLTDNNG-NVSLAQVR 382  
DB 366 QSANNPTVKVKHIGSDLAESVYAKLD--DTSKFEDVTEKVNLSYTSNGGYTLNLGLD 423

QY 383 N-DYIV-FEGD-----GY-----YYQTYDQWKATNKA---DNIAM 413  
DB 424 NSKDYIKYEGEYDONAKDLNFRTHLSGHHKYYPYPPYPPVQLTWNGVAFYSNNAK 483

QY 414 RDAH-----VIEDGNGDRLVFEASTGLENYQ- 441  
DB 484 GDGKDKPNDPITIEKSEPIDLDIKSEPPVEKHELTGTTEESNDSKPIDFHYHTAVEAGEH 543

QY 442 -----EDQIY-----NWLNYGGDDA-----FNISLFR 464  
DB 544 AEGIIETEDSIHVDFFEESTHENSKKHADYVEEEDTNPGGQVQVTTESNLVEDEESTKG 603

QY 465 ILSN-----DDIKSRATWANAAGILKLNKDEKNPKVAELYSPLISAPVSDIERP 516  
DB 465 ILSN-----DDIKSRATWANAAGILKLNKDEKNPKVAELYSPLISAPVSDIERP 516

Db	604	IVTGVSDHTTVBTDKREYTTESN	-----LIEL-VDELPHEGQAQGP	-----EETEN	651	
Qy	517	N	-----VWKLGNKYFLFAATRLNRGSDDAWNNYAVGDVNAVGYVADSLTGSYK		568	
Db	652	NHHISHSGLTENGHNY	-GVDEIEENSVDIKSELGYEGQSGNQSFED	--TEBDK	708	
Qy	569	PLNDSGVLTASVPANWRATYSYXAVP	-VAGKDDQVLVTSYMTNRNGVAGKGMDSW-A		626	
Db	709	PXYEOG	-----GNIVDIDFSVPIHQCN	-----NGNQSFEDTEEDK	746	
Qy	627	PSF	-----LLQINPDNTTVLAKMTNCGDWIDDSSE	--NLDWIG	-----DLDSAALP	672
Db	747	PXYEOGGNIIDIDFSVPIHG	-FNKHNEIIEEDTKDKPNYQFGHNSVDFEEDTLPKV		805	
Qy	673	---	GERDPVDWDLIGYGLKPHDPATPNDPETPTTPTPTPTPTNP	--KTPKTPENPGTPQ	727	
Db	806	SGONEGQOTIEDTT	--PPTPTPEVSEPEPTP	-TPPTPEVSEPEPTP	TPKPEVSEPE	862
Qy	728	TPNTPTPEIPLTP	-----ETPKOP	-----	ETOT	751
Db	863	TP-VPTPEVSEPKVPVPAKEPKPKSPV	VEQKVVTPVIEINEKVKAVAPTQKQSK		921	
Qy	752	NNRLPQTG	--NNANKAMIGLGMGTLLSMFGLAEINKRREN		789	
Db	922	KSELPETGGEESTNKGML	-----FGGLF\$ILGLVLLRRKN		958	
RESULT	14					
ID	Q9KFW0					
AC	Q9KFW0	PRELIMINARY;	PRT;	1661	AA.	
DT	01-OCT-2000	(TREMBlrel. 15, Created)				
DT	01-JUN-2000	(TREMBlrel. 15, Last sequence update)				
DT	01-JUN-2001	(TREMBlrel. 17, Last annotation update)				
DE	BH0361	PROTEIN.				
GN	BH0361.					
QS	Bacillus halodurans.					
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;					
OC	Bacillus/Staphylococcus group; Bacillus.					
OX	NCBI_TaxID=86665;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-C-125 / JCM 9153;					
RX	MEDLINE=20512582; Pubmed=11058132;					
RA	Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,					
RA	Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,					
RA	Horikoshi K.					
RT	Complete genome sequence of the alkaliphilic bacterium Bacillus					
RT	halodurans and genomic sequence comparison with Bacillus subtilis.;					
RL	Nucleic Acids Res. 28:4317-4331(2000).					
DR	EMBL; AP001508; BAB04080.1;					
DR	InterPro; IPRO01899; Gram_pos_anchor.					
DR	Pfam; PF00746; Gram_pos_anchor; 1.					
DR	PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.					
DR	Complete proteome.					
Qs	SEQUENCE	1661	AA.	183657	MW; 837658930D2E60C4 CRC64;	

Query Match	5.68;	Score 230;	DB 16;	Length 1661;
Best Local Similarity	20.48;	Pred. No. 0.013;		
Matches 181;	Conservative 110;	Mismatches 288;	Indels 310;	Gaps 41;
Qy	92	VASSAAITSSSTAASLNTDSKAAQENTAKNDOTQRAANESSEAKNPAYNVNDS	151	
Db	885	IDSAYWTFKT-----SLEDTLIKSNYPNRATLKSD-----	111	
Qy	152	SAAKNDQOQSSKNNTAKLNKDAENVYVKAGIDPNSLTDQIKALKNMFSKAASK---	208	
Db	926	VSVRHGE-----FVTKSGAONGRLDWSL-----NINFSQSISNAKI	964	
Qy	209	-----TQMTYNDFOQTADFLIKDGRYTVPFKASEI-----KNMPAATTKQAQNTWTEPL	259	
Db	965	VDEPSVNGILLEDTFRLYATNVQPGTWT-----KGDELERDKDYTLDTRTDGEGNQHFL	1020	

QY 260 DVWDSNPVQDVRTQGVANNNGYQLVITAMGIPNDNDNHILLYKNYGDNELSHKKNVGP 319  
Db 1021 QF-----TEDIDTGYILEYQSF--INARG--DVVTNNVKGEDQLTTEEYTE-----1064  
QY 320 FGYNSTAVSGWSSGAVLNSDINSIOLFTRVSDNTNHNOKIASATLYLDNNGNVSLA 379  
Db 1065 ---SSGAVSVRLSGAGTSGTSGLEVTKYDAUTG---EVLQATFTLYDSGSEFAIR 1117  
QY 380 QVR-----NDYIVFEG---DGYIYOTYDOWKATNKAGDNIAMRDH--VIE 420  
Db 1118 TLETGEDGKATFVNLLYGYLLKEDSAPGYLVINGDTQVTT---IDTV---LHEVTVE 1170  
QY 421 DNGDRYLV-----FEASTGLEWYQGEDQIYNLWYGGDDAFENIKSLFRILSND 470  
Db 1171 NEKSDINRVSAGVQVKVDEETG--ESLQ-----ALFALQOKVD 1210  
QY 471 IKSRTAWANAIGILKNKDEK-----NPKVAELYSP-----LISAPMV--SDIE 514  
Db 1211 -----DEFVTTAEMETDEBEGIVFAGSLPBGDYQFVELNAPGVKDETPVVFVEED 1262  
QY 515 RPNVVKLGKNYLF--AATRLNRGSDADMANVAVG-----DNYAMVGYVADSLT 564  
Db 1263 RTEIQLKENHLLPGSVQLVKVDADA---ANTLEGAFTLLDGGENVQEGLTID---1316  
QY 565 GSYKPLNDSGVVLTASVPANRR-----TATYSYVAVPV-----597  
Db 1317 -----ENGVOVVYTLKPGEYQFVETKAPAGYELATPGFTIERNQQEVATVAVENHLI 1370  
QY 598 -----AGKDD-----QVLVTS-----YMTNR 613  
Db 1371 PGSVQLVKVDADDAATLBCAFESLLDGGENVQEGLTATDAGVVVTLKPGEQFVET 1430  
QY 614 NGVAGKGMSTWAPSLLOINPDNTTV-----LAKMTNCGDWIWDSSSE--N 659  
Db 1431 KAPAGYELDAT--PIGTETIERNQQEVATVAVENHLIPGVSQVLKVDREDSATVLEGAESL 1489  
QY 660 LDMIGD--LDSAALPGERD-----KPVMDWLI-----GY-----686  
Db 1490 LDEEGNVLREGIRTGENGQMLVIDLKPGDYQFIETKATPGYELDDTPIAFTIERGTDEV 1549  
QY 687 -----GLKPHDPATPNDEPTTTPETPTNPKTPKTPENPGTPTPTNPTNPEPLT 740  
Db 1550 TLVFNKLTPEVPSPENPENPENPENPEIPELTPEPENPETPESPRHSEDEVPNS 1609  
QY 741 PETPKQPETQTN--NRLPQTGNNAKAMIGLGMGLTLLSMGLAENKRR 787  
Db 1610 DKNPEKSSDNNRIDRLPOTGEEFTLLIIGL--LMTAGGVLLLLKRR 1636  
  
RESULT 15  
ID Q9LQA7 PRELIMINARY; PRT; 138 AA.  
Q9LQA7  
AC Q9LQA7;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE FAN2.10.  
CS Arabidopsis thaliana (mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids I; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_Taxid:3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Shinn P., Brooks S., Buehler E., Chao Q., Dunn P., Khan S., Kim C.,  
RA Altafi H., Araujo R., Conn L., Conway A.B., Gonzalez A., Hansen N.F.,  
RA Huizar L., Kremenetskaia I., Lenz C., Li J., Liu S., Luros S.,  
RA Rowley D., Schwartz J., Toriumi M., Vysotskaia V., Yu G., Davis R.W.,  
RA Federspiel N.A., Theologis A., Ecker J.R.,  
RT "Genomic sequence for Arabidopsis thaliana BAC FAN2 from chromosome  
RT I.";  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.

	Query Match	5.4%	Score 224;	DB 10;	Length 138;	
	Best Local Similarity	60.9%	Pred. No. 0.0015;	Mismatches	0; Indels	0; Gaps
	Matches	42;	Conservative	0;		
Qy	690 PHDPATPNDPTTTPETPPNPNTKPKTKPNPGTPQTENPTNTEPEILTTETPKQPET	749				
Dd	33 PGIPATPATPATPATPATPATPTPTFTPTSIPAIPATPAITPLTETTTTTTT	92				
Qy	750 QTNNRLPQT	758				
Dd	93 PTTPLTPPT	101				

Search completed: September 26, 2002, 18:27:55  
Job time: 254 sec





```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model
Run on:      September 26, 2002, 18:23:41 ; Search time 22.4 Seconds
              (without alignments)
              1363.827 Million cell updates/sec

Title:
Perfect score:  US-09-995-587A-1
Sequence:       4129
                1 MYKSKNWAVTLSTAALVF.....GMGTLLSMFGLAEINKRRFN 789

Scoring table:
                BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:      105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters:      105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :      SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

```

## SUMMARIES

Result No.	Query Match %	Score	Length	DB	ID	Description
1	2016.5	48.8	797	1	SACB_STRMU	pl1701 streptococc
2	1739.5	42.1	969	1	SACB_STRSL	Q55242 streptococc
3	744	18.0	473	1	SACB_BACSU	Q05655 bacillus su
4	741.5	18.0	472	1	SACB_BACAM	P21130 bacillus am
5	721	17.5	473	1	SACB_BACST	P94468 bacillus st
6	304	7.4	423	1	SACB_ZYMKO	Q60114 zymomonas m
7	301.5	7.3	413	1	INVB_ZYMKO	Q60115 zymomonas m
8	281	6.8	415	1	SACB_ERWAM	Q46654 erwina amy
9	275.5	6.7	415	1	SACB_ERHAG	Q54435 rhnella ag
10	274	6.6	431	1	SACB_PRESH	Q68609 pseudomonas
11	266	6.4	584	1	SACB_ACEDI	Q43998 acetobacter
12	257	6.2	415	1	SACB_PSESG	Q52408 pseudomonas
13	253	6.2	1018	1	FNBA_STAAU	P14738 staphylococ
14	241	5.8	1164	1	BAG_STRAG	P27951 streptococc
15	229.5	5.6	1500	1	SSP5_STPRGN	P16952 streptococc
16	227	5.5	1337	1	DEXT_STRDO	P39653 streptococc
17	220	5.3	826	1	SSP2_PLAYO	Q01443 plasmodium
18	213	15.2	1565	1	PAC_STRMU	P11657 streptococc
19	211.5	5.1	890	1	DEXT_STRMU	Q54443 streptococc
20	208.5	5.0	1256	1	MRP_STRSU	P32653 streptococc
21	205.5	5.0	1637	1	MRSP_STAAU	P80544 streptococc
22	199	4.8	1592	1	GTF2_STRDO	P27470 streptococc
23	195	4.7	1758	1	YIR7_YEAST	P40434 saccharomyc
24	195	4.7	1758	1	YIW5_YEAST	P40889 saccharomyc
25	191.5	4.6	1561	1	SNAP_STRMU	P23504 streptococc
26	190	4.6	1597	1	GTF1_STRDO	P11001 streptococc
27	185.5	4.5	1528	1	SPAA_STRDO	P21979 streptococc
28	185	4.5	1569	1	YFJA_ECOLI	P52143 escherichia
29	184	4.5	1462	1	GTFD_YEAST	P49331 streptococc
30	183.5	4.4	881	1	YVH8_YEAST	P47033 saccharomyc
31	182.5	4.4	1694	1	IGA0_HAETI	P44969 haemophilus
32	182.5	4.4	1296	1	ASAL_ENTFA	P17953 enterococcu
33	182.5	4.4	1396	1	VITE_BP75	P13390 bacteriacea

**F13330 bacteriophage**

34	182	4.4	995	1	YIO9_YEAST
35	182	4.4	1296	1	VAC1_HELPY
36	181.5	4.4	1423	1	FRUA_STROM
37	179.5	4.3	1287	1	VAC2_HELPY
38	179.5	4.3	1702	1	IGA2_HAEIN
39	178.5	4.3	1120	1	STFR_ECOLI
40	173	4.2	1365	1	GTF5_STROD
41	172.5	4.2	1310	1	VAC3_HELPY
42	172	4.2	1250	1	YEAL_ECOLI
43	171.5	4.2	857	1	NFM_CHICK
44	170	4.1	1183	1	CNA_STAAR
45	169.5	4.1	1827	1	MAP2_HUMAN

## ALIGNMENTS

```

RESULT 1
SACB_STRMU
ID SACB_STRMU STANDARD; PRT; 797 AA.
AC P11701;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Levanucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl
DE transferase) (Sucrose 6-fructosyl transferase).
GN FTF.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=88115184; PubMed=2828325;
RA Shiroza T., Kuramitsu H.K.;
RT "Sequence analysis of the Streptococcus mutans fructosyltransferase
RT gene and flanking regions".
RL J. Bacteriol. 170:810-816(1988).
CC -! CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL)(N) =
CC GLUCOSE + (2,6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN
CC ALSO ACT AS FRUCTOSYL ACCEPTORS).
CC -! SUBCELLULAR LOCATION: Secreted.
CC -! SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
EMBL; M18954; AAA88584.1; --
DR PIR; B28551; B28551.
DR InterPro; IPR003469; Glyco_hydro.68.
DR Pfam; PF02435; Glyco_hydro.68; 1.
DR Transferase; Glycosyltransferase; Signal.
FT SIGNAL 1 34
FT CHAIN 35 797
FT LEVANSUCRASE.
SQ SEQUENCE 797 AA; 87574 MW; 6DD057465DB4042 CRC64;

Query Match 48.8%; Score 2016.5; DB 1; Length 797;
Best Local Similarity 50.4%; Pred. No. 2.6e-93;
Matches 408; Conservative 118; Mismatches 211; Indels 73; Gaps 14;

QY 1 MYKSGKNWAVVLSTAAVL-FGATTVNASADTNIENNDSSTVQVTTGDDNDIAKSVTLGS 59
Db ||||| 1 1:|||| 1:|||| 1:|||| 1:|||| 1:|||| 1:|||| 1:|||| 1:|||| 1:
9 MYKKGFWVAITITAMLFGIGLSVQA-----DEANSTQV-----S 45
QY 60 GOVSAASDITITRTSANANASSAANTONSNOVASSAITSTSSAASLNNTD--SKAAQ 117
::: 1 1:||||: 1:|||| 1:|||| 1:|||| 1:|||| 1:|||| 1:|||| 1:|||| 1:

```

RESULT	2	SACB_STRSL	STANDARD;	PRT;	969 AA.
ID		SACB_STRSL			
AC	Q55242;				
DT	15-DEC-1998 (Rel. 37, Created)				
DT	15-DEC-1998 (Rel. 37, Last sequence update)				
DT	15-DEC-1998 (Rel. 37, Last annotation update)				
DE	Levananase precursor (PC.2.4.1.10) (Beta-D-fructofuranosyl				
DE	transferase) (Sucrose 6-fructosyl transferase).				
GN	FTF.				
OS	Streptococcus salivarius.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;				
QC	Streptococcus.				
OX	NCBI_taxid-1304;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RP	SEQUENCE ATCC 25975;				
RC	MEDLINE-93322332; PubMed-8331080;				
RC	Rathsam C., Giffard P.M., Jacques N.A.;				
RT	"The cell-bound fructosyltransferase of Streptococcus salivarius: the				
RT	carboxyl terminus specifies attachment in a Streptococcus gordonii				
RT	model system".				
RE					

[illegible]

Tang L.B., Lenstra R., Borc

Isolation and characterization of levansucrase-encoding gene from *Levanderia* sp. *Levanderia* N.7

RESULT 5  
SACB\_BACST  
ID SACB\_BACST  
AC P94468: STANDARD; PRT; 473 AA.

```

Query Match      17.5%;   Score 721;   DB 1;   Length 473;
Best Local Similarity 36.3%;   Pred. No. 3.3e-29;
Matches 181;   Conservative 90;   Mismatches 149;   Indels 78;   Gaps 18;

QY 199 MNFSAAKSGTQMY-----NDFOKIADTLIK 225
Db 1 MNIKFAQAQVLVFTTALLAGGATQAFAKETNOKPKYETGISHTRDMLOIPEQ--Q 58
QY 226 QDGRVYVFFKASIKNNPAATTKDAQTNTIEPLDVWDSWPQDVRTGVANNGYQLVI 285
Db 59 KNEKYQVPEFDSITKNISSA-----KGLDVWDSWPLONA-DGTVAHYGHYTFV 107
QY 286 AMWGIP-NQNDNHVILLYNKYGDNLSHWNKVGPI-----EGYNSTAV--SOWSGSA 335
Db 108 ALAGDPKNADDTSYMFYQKVGTSIDSWKTFGRYFKDSKDFDANDSLTKDQTQWSSGA 167
QY 336 VLNSDSIQLFYRVVTSNNTNHQKIASATLYLTDDNNGVNSLAQVRNDYLVFPGDYYY 395
Db 168 TFTSDGKIRUFYF--DFSGKHYGKQTLTAQVNVSSADSSLINGVEDYKSFIDGDSKT 225
QY 396 QTYDOW--KATNKGADNIAMRDAAHVIEGNGDRYLVFEASTGLN--YGEDQIYNMNLNYG 452
Db 226 QNVQOQFIDEGNSYSGDNHTLRDPHYVED-KGHKLVLFEAFTGTEDGYQGSESLFNKAYG 284
QY 453 GDDAFNIKSLFRLNSDDIKSR-ATWANAAGILKLNKDEKNPKVAELYSPLLSAPWSD 511
Db 285 KSTSFPRQSQKLLQSD--KNRTAELANGALGMTELNDVYTLKKYMK---PLASTVTD 339
QY 512 ETERPNNVKLGNKYLYFAATRLNRGNSDDAMNNANYAGDVNAMGVYVADSLTGSYKPLN 571
Db 340 ETERANVFKNMGKWL-----STDSRGSQ-----MTIDGTSNDIIMLGIVNSLTPGYKPLN 392

```

QY 572 DSGVLTASYPANRTATSYAVPVAGKDDQVLVTSTNTRNGVAGKGMDSWAPSFL 631  
 Db 393 KTLGLVKMDLDNDVTFYSHFAPQA-TGNVVTSTNTRNGFYADK--QSTFAPSFL 449  
 QY 632 QINPONTTTLAKMTNOG 649  
 Db 450 NIOGKKTSVKASILDQ 467

RESULT 6  
 SACS\_ZYMMO  
 ID SACS\_ZYMMO STANDARD; PRT; 423 AA.  
 AC Q60114; Q60116; Q06487;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Levanucrase (EC 2.4.1.10) (Beta-D-fructofuranosyl transferase)  
 DE (Sucrose 6-fructosyl transferase).  
 GN SACS OR LEVU OR SUCE2.  
 OS Zymomonas mobilis.  
 OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;  
 OC Zymomonas.  
 OX NCBI\_TaxID=542;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 10988 / ZM1;  
 RX MEDLINE-93305726; PubMed-8318541;  
 RA Song K.B., Joo H.K., Rhee S.-K.;  
 RT "Nucleotide sequence of levansucrase gene (levu) of Zymomonas mobilis  
 RL extracellular levansucrase and invertase.";  
 RN Biosci. Biotechnol. Biochem. 59:289-293(1995).  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NRRL B806;  
 RA Gunasekaran P., Mukundan G., Kannan R., Velmurugan S.,  
 RL Alt-Abdelkader N., Alvarez E., Baratti J.;  
 RP Submitted (JUN-1994) to the EMBL/Genbank/DBJ databases.  
 RP [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 31821 / ZM4 / CP4;  
 RA Ahn J.Y., Kang H.S.;  
 RP "Sequence analysis of 4486 fosmid clone of Zymomonas mobilis ZM4.";  
 RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL)(N) =  
 CC GLUCOSE + (2,6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN  
 CC ALSO ACT AS FRUCTOSYL ACCEPTORS).  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- PTM: DOES NOT SEEM TO BE N-TERMINALLY PROCESSED.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; AF081588; AAA27695.1;  
 DR EMBL; D17524; BAA04475.1;  
 DR EMBL; L33402; AAA27702.1;  
 DR EMBL; AF313764; AAC29870.1;  
 DR InterPro: IPR003469; Glyco\_hydro\_68.  
 DR Pfam: PF02435; Glyco\_hydro\_68; 1.  
 KW Transferase; Glycosyltransferase.

FT CONFLICT 39 39  
 FT CONFLICT 200 203  
 FT CONFLICT 217 217  
 FT CONFLICT 220 220  
 FT CONFLICT 244 244  
 FT CONFLICT 379 379  
 SQ SEQUENCE 423 AA; 46762 MW; 19A4691DA3EB9FE3 CRC64;

Query Match  
 Best Local Similarity 7.4%; Score 304; DB 1; Length 423;  
 Matches 120; Conservative 74; Mismatches 155; Indels 102; Gaps 23;

QY 250 DAQNTNTEPLD-----VWDSHPVDVTRTGOVANNWGYQLVIAWM-----GIN 292  
 Db 24 DDPATMTIDYDFPVMTDKYVWDTWPLRDI-NGOVVSFGMSVIFALVADRTKYGWHN 82  
 QY 293 QNDN-HIYLLNKGDNELSHKKNVGFPGYNSTAVSOWSSGSAVL--NSDNSIOLFTR 349  
 Db 83 RNDGARIGYFSRGGSN-----WIFGGHLLKDGANPRSWSGCTINAPGTANSVEVFTS 138  
 QY 350 VDTSDNNTNHQI-ASATLYLTDNNGVSLAQVRNDYIVFEGDGYIYQTYDQWKATNKA 408  
 Db 139 V-----NDTPSEVPACQKGIYADOKSWFDFKVTDLFOADGLYADY----AENFW 190  
 QY 409 DNTAMRDAHV-TEGNGDRYLVFEASTGLENYQGEDQIYNWLNLYGGDDAFNIKSLFRILS 467  
 Db 191 D---FRDPHFVFNPDGKTYALFEGVAMER-----GTAVAGEEIGPVPP 233  
 QY 468 NDDIKSRATWANAAGILK-LNKDEKNKVAELYSPLISAPMVSEIERPNWVKLGNKY 526  
 Db 234 KTETPDGARYCAAGIAQALNEATEWK---LLPPLVAFGVNDQTERPHVVFQGLTY 290  
 QY 527 LFAAFLNNGSNDAMNANYAVG---DNVAMVGYADS-LTGSYKPLNDGYYVLTASV 581  
 Db 291 LFTISH-----HSTYADGLSGPDGV--YGFSENGIFGPEPLNGSLGLV--GN 335  
 QY 582 PANWRATATSYAVPVAGKDDQVLVTSTN-----RNVAGKGMDSWAPSFLQ 632  
 Db 336 PSSQPYQAYSHVM-----TNGLVTSFIDTIPSSDPNVTRYG-----GTLAPTIKLE 382  
 QY 633 INPDFTTTLAK-----MTNQGDWIWDDSEN 659  
 Db 383 LVGHRSFTEVKGYGYPQIENLAEDSSN 413

RESULT 7  
 INVB\_ZYMMO  
 ID INVB\_ZYMMO STANDARD; PRT; 413 AA.  
 AC Q60115; Q60117; Q60125;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Extracellular sucrose (EC 3.2.1.26) (Beta-fructofuranosidase)  
 DE (Invertase) (Saccharase) (Protein B46).  
 GN SACC OR INVB OR SUCE3.  
 OS Zymomonas mobilis.  
 OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;  
 OC Zymomonas.  
 OX NCBI\_TaxID=542;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN-ATCC 10988 / ZM1;  
 RX MEDLINE-94368848; PubMed-8086457;  
 RA Song K.B., Lee S.K., Joo H.K., Rhee S.-K.;  
 RT "Nucleotide and derived amino acid sequences of an extracellular  
 RL sucrose gene (invb) of Zymomonas mobilis ZM1 (ATCC10988).";  
 RN Biochim. Biophys. Acta 1219:163-166(1994).  
 RP [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-IFO 13756;  
 RX MEDLINE-95218269; PubMed-7766026;  
 RA Kyono K., Yanase H., Tonomura K., Kawasaki H., Sakai T.;

RT "Cloning and characterization of *Zymomonas mobilis* genes encoding  
 RL extracellular levansucrase and invertase";  
 RN Biosci. Biotechnol. Biochem. 59:289-293(1995).  
 RP [3]  
 RC SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX STRAIN-NREL B806;  
 RA MEDLINE-95297907; PubMed-7778976;  
 RA Kannan R., Mukundan G., Ait-Abdelkader N., Augier-Magro V.,  
 RA Baratti J., Gunasekaran P.;  
 RT "Molecular cloning and characterization of the extracellular sucrose  
 RL gene (sacC) of *Zymomonas mobilis*";  
 RN Arch. Microbiol. 163:195-204(1995).  
 RP [4]  
 RC SEQUENCE FROM N.A.  
 RA STRAIN-ATCC 31821 / ZM4 / CP4;  
 RA Ahn J.Y., Kang H.S.;  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 CC -|- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-  
 CC fructofuranoside residues in beta-D-fructofuranosides.  
 CC -|- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AF081588; AAC61488.1;  
 DR EMBL; D17524; BAA04476.1;  
 DR EMBL; L33403; AAC36942.1;  
 DR EMBL; AF313764; AAG29871.1;  
 DR InterPro; IPR003469; Glyco\_hydro\_68.  
 DR Pfam; PF02435; Glyco\_hydro\_68; 1.  
 KW Hydrolase; Glycosidase.  
 FT CONFLICT 85 85 G -> A (IN REF. 3).  
 FT CONFLICT 105 105 S -> L (IN REF. 2).  
 FT CONFLICT 108 108 T -> S (IN REF. 2).  
 FT CONFLICT 404 407 PYWP -> LGMA (IN REF. 3).  
 SQ SEQUENCE 413 AA; 46100 MW; BACC12D167332E47 CRC64;

Query Match  
 Best Local Similarity 7.3%; Score 301.5; DB 1; Length 413;  
 Matches 108; Conservative 79; Mismatches 141; Indels 75; Gaps 23;  
 QY 257 EPLDVNDSPVQDVRTGQVA--NMNGYQVIAAMGIPNDNH---IYLLYKYGDNEL 310  
 DB 38 DDLALMDTWPLRDINGNPVSKGNVIFSLVADRNP-WNDRHSHARIGYFSKOGKS-- 94  
 QY 311 SHWKNVGPFGYNSTAVSQWSSGSAVL--NSDNIQLFYTRVDTSDNNTNKHIAAT-- 366  
 DB 95 --WYGGHLLQESANTRAEASGGTMAPGRNQVETFTST-LFDKNGVREAAVATKG 151  
 QY 367 -LYLDNNGNYSIAQVRNDYIVFGDGYGYQYTDQKATNKGADNIAADAHV-IEDGNG 424  
 DB 152 RIY-ADSEG-VWTFKGFQSDTLDFQADGLFYQNYAENLNN-----FRDPVFVIMPEDG 202  
 QY 425 DRYLVFEASTGLNYQGEDQIYNWLNNGYQVIAAMGIPNDNH---IYLLYKYGDNEL 310  
 DB 203 ETYALFEAN--VATVRGEDI-----GED-----EIGPVPVATVPKDNLCASIGI 248  
 QY 485 LKLNKDEKNPKVAELYSPLASPMVSDIEPRNVKLGKYYLEFAATRLNRSNDNIAAMR 414  
 DB 249 ARCLSPORTE--WELLPLLLTARFVNDQMERPHVIFQNGLTLYLFISH-----D 295  
 QY 545 ANYAVG--DNVAMVGVAQDS-LTGSKVPLDNGSVLTVASVPANMTATSYAVPVACKG 601  
 DB 296 STYADGLTGDGLGVFSENGIFGPEPLNGSLVLGG--PASOPTAYAHYIM----- 347  
 QY 602 DQVLVTYMTN-----RNG--VAGKGMDSWAPSEFLQINPDNT 638

Db 348 NNGLVESFINEIDPKSGKVIAGGSL-----APTVRVELOGHET 386

RESULT 8  
 SACS\_ERWAM  
 ID SACS\_ERWAM STANDARD; PRT; 415 AA.  
 AC Q46654;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Levansucrase (EC 2.4.1.10) (Beta-D-fructofuranosyl transferase)  
 DE (Sucrose 6-fructosyl transferase).  
 GN LSC.  
 OS Erwinia amylovora.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Erwinia.  
 OX NCBI\_TaxID=552;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EA7/74;  
 RA Geier G., Gelder K.K.;  
 RT "Characterization and influence on virulence of the levansucrase gene  
 RT from the fireblight pathogen *Erwinia amylovora*.";  
 RL Physiol. Mol. Plant Pathol. 42:387-404(1993).  
 CC -|- CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL) (N) -  
 CC GLUCOSE + (2,6-BETA-D-FRUCTOSYL) (N+1) (OTHER SUGARS CAN  
 CC ALSO ACT AS FRUCTOSYL ACCEPTORS).  
 CC -|- SUBCELLULAR LOCATION: Secreted.  
 CC -|- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; X75079; CAA52972.1;  
 DR InterPro; IPR003469; Glyco\_hydro\_68.  
 DR Pfam; PF02435; Glyco\_hydro\_68; 1.  
 KW Transferase; Glycosyltransferase.  
 SQ SEQUENCE 415 AA; 46408 MW; 4FF564F6E0607FEB CRC64;

Query Match  
 Best Local Similarity 6.8%; Score 281; DB 1; Length 415;  
 Matches 109; Conservative 73; Mismatches 137; Indels 106; Gaps 22;  
 QY 257 EPLDVNDSPVQDVRTGQVA--NMNGYQVIAAMGIPN-----QNDNHIYLL---YNYKG 306  
 DB 40 ERFVFWTMDPLRDF-DGEITSVNCWCIIFTADRNTPNQFDENGNDITRDWEDRHG 98  
 QY 307 DNELSHW-KNVGP--IFGYNSTA-----VSQWSSGSAV-LNSDNIQLFYTRVDTSDNNT 357  
 DB 99 RARICYWTSRTGKDWIFGGRVMAEGVAPTREWAGTPIILLNDRGIDIDLYTCV-----T 152  
 QY 358 NHQIASATLYLTNDNNGNYSIAQVRNDYIVFGDGYGYQYTDQ---WKATNKGADNIAAMR 414  
 DB 153 PGATIAKVRGKIYTSQSVSLEGFQVTSLSADGTIYQTEQNAFN-----FR 202  
 QY 415 D-AHVIEDNGDRYLVEASTGLNYQGEDQIY-----NWLNYGDDAFNKSIFR 464  
 DB 203 DPFPFIDRNDGKLYMLFEGN--VAGPRGSHEITQAEWNPVPGYEDVGG----- 249  
 QY 465 ILSNDDIKSRATWANAAIGILKLNKDEKNPKVAELYSPLASPMVSDIEPRNVKLGK 524  
 DB 250 -----AKYQACGVG-LAVAKDLGSGE-WQILPLPLITAVGVNDQTERPHEVFDGK 297  
 QY 525 YYLFAATRLNRSNDNIAAMNANYAVGDNY-----AMVGVAQDSLTGSKVPLDNGSVLTVAS 580  
 DB 298 YLFT-----ISHKYTFADNLGTGDYGVFVSKLGTPTPMNSGLVL--G 342

QY 581 VPANRRATATSYAVPVAGKDDQVLTYSYMTNRRGVAGKMD-----STWAPSFLLQINPD 636  
 Db 343 NPSSQPPQTYSHYMP-----NGLVTSFI-----DSVPKKGKDYRGIGTEAPTCKILLKGD 393  
 QY 637 NTYTV 641  
 Db 394 RSFIV 398  
 RESULT 9  
 SABC\_RAHQ  
 ID SABC\_RAHQ STANDARD; PRT; 415 AA.  
 AC O54435;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Levanucrase (EC 2.4.1.10) (Beta-D-fructofuranosyl transferase)  
 DE (Sucrose 6-fructosyl transferase).  
 GN SABC.  
 OS Rahmella aquatilis.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Rahmella.  
 OX NCBI\_TaxID=34038;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 33071;  
 RX MEDLINE=99127094; PubMed=9928133;  
 RA Song K.B., Seo J.W., Kim M.G., Rhee S.K.;  
 RT "Levanucrase of Rahmella aquatilis ATCC33071. Gene cloning,  
 RT expression, and levan formation.";  
 RL Ann. N.Y. Acad. Sci. 864:506-511(1998).  
 CC -/- CATALYTIC ACTIVITY: SUCROSE + (2.6-BETA-D-FRUCTOSYL)(N) =  
 CC GLUCOSE + (2.6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN  
 CC ALSO ACT AS FRUCTOSYL ACCEPTORS).  
 CC -/- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC EMBL; U91484; AAC36458.1;  
 DR InterPro: IPR003469; Glyco\_hydro\_68.  
 DR Pfam: PF02435; Glyco\_hydro\_68; 1.  
 KW Transferase; Glycosyltransferase.  
 SQ SEQUENCE 415 AA; 45939 MW; E028828813D13A74 CRC64;

Query Match 6.7%; Score 275.5; DB 1; Length 415;  
 Best Local Similarity 27.2%; Pred. No. 3.9e-07;  
 Matches 104; Conservative 58; Mismatches 132; Indels 89; Gaps 19;  
 QY 261 WDSWPVQDVRGQVAVNGVQLVAMGIPNQNDHIYLLN-----LYNKYGDNE 309  
 Db 44 INDTPMLRSL-DGTVSVVDSWVITLTAQRN-NNSEYLDAGNYDITSDNNRHRAR 101  
 QY 310 LSHWKN-----VGPIFGYNSTAVSQWGSVAV-LNSDNSIQLFYTRVDTSDNTHNQ 101  
 Db 102 ICYVYVSRGKDMIFGRVVAEGVSPTSREWAGTPIILLNEDGIDLYTCV-----TPGA 155  
 QY 361 KTASATLYLTDNNGVSLAQVRNDYIVFEGDGYVYQYDQ-----WKATNKGADNIAMRD-A 416  
 Db 156 TIARVGRKVLTSSEGVTLAGNEVKSIFSDGVYVYQESQNPYN-----FRDPS 205  
 QY 417 HVIEDNGDRYLVEASTGLENYOGEDQIYNWLNNGYGGDDAFNKSFLRILNSNDIKSRAT 476  
 Db 206 PFIDPHDKLTWFEVGNVAGE--RGSHVI-----GQENGTLPPGHROVGN-----AR 251  
 QY 477 WANRAIGL-----KLNKDEKPKVAELYSLPISAPMYSDEIERNVVYKLGNYLLFAATRL 533

Db 252 YQAGCTGMAVAKDLSGE-----WEILPLVTAAGVNDQTERPHFVFDGKYLLFTISH- 305  
 QY 534 NRGSDNDAMNANYAVG-----DNVAMGVVADSLTGSYKPLNDGSLVLTASVPANWRTAT 589  
 Db 306 -----KFTYADGLTGPQGV--YGFLSDNLTLGTPSPMNGSLVL--GNPSPQFPQT 351  
 QY 590 YSYIYAVPVAGKDDQVLTYSYMTN 612  
 Db 352 YSHCVMP-----NGLVTSFIDN 368  
 RESULT 10  
 SABC\_PSESH  
 ID SABC\_PSESH STANDARD; PRT; 431 AA.  
 AC O68609;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Levanucrase (EC 2.4.1.10) (Beta-D-fructofuranosyl transferase)  
 DE (Sucrose 6-fructosyl transferase).  
 GN LSC.  
 OS Pseudomonas syringae (pv. phaseolicola).  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas  
 OX NCBI\_TaxID=319;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCPPB 1321;  
 RX MEDLINE=98394981; PubMed=9726857;  
 RA Hettwer U., Jaekel F.R., Boch J., Meyer M., Rudolph K., Ullrich M.S.;  
 RT "Cloning, nucleotide sequence, and expression in Escherichia coli of  
 RT levanucrase genes from the plant pathogens Pseudomonas syringae pv.  
 RT glycolina and P. syringae pv. phaseolicola.";  
 RL Appl. Environ. Microbiol. 64:3180-3187(1998).  
 CC -/- CATALYTIC ACTIVITY: SUCROSE + (2.6-BETA-D-FRUCTOSYL)(N) =  
 CC GLUCOSE + (2.6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN  
 CC ALSO ACT AS FRUCTOSYL ACCEPTORS).  
 CC -/- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC EMBL; AF052289; AAC36063.1;  
 DR InterPro: IPR003469; Glyco\_hydro\_68.  
 DR Pfam: PF02435; Glyco\_hydro\_68; 1.  
 KW Transferase; Glycosyltransferase.  
 SQ SEQUENCE 431 AA; 47603 MW; 795FEA246D40C40F CRC64;

Query Match 6.6%; Score 274; DB 1; Length 431;  
 Best Local Similarity 25.1%; Pred. No. 4.8e-07;  
 Matches 115; Conservative 66; Mismatches 165; Indels 112; Gaps 21;  
 QY 238 SEIKNMP-----AATTKDAQNTIEPL-----DYKDSMPVQD 269  
 Db 9 SOLKSLAGINIEPTVMSRADALKVNEPDPTTQPLVSADFPVMSDVTYFINDTMPLRE 68  
 QY 270 VRTGQVANNNGVQLVAMGIPNQNDHIYLLN-----KYGDNELSHWKN--- 315  
 Db 69 L-DGTVSVVDSWVITLTAQRNHPNDQ-YLDANGRYDIKRDWEDRGRARMSYVTSRTG 126  
 QY 316 -----VGPIFGYNSTAVSQWGSVAV-LNSDNSIQLFYTRVDTSDNTHNQIASATLYL 369  
 Db 127 KDWTFGRVMAEGVSPTSREWAGTPIILLNEDGIDLYTCV-----TPGAAIARVGR 180  
 QY 370 TDNNGVSLAQVRNDYIVFEGDGYVYQYDQWKATNKGADNIAMRD-AHVIEDNGDRYL 428



```

Db 181 VTSQGVLEKDFQVKKLFADGTYTYQTEAQNSSWN-----FRDPSPFDPNDGKLYM 233
QY 429 VFEASTGLENYQGDQIYNNLYNMGDDAFNIKSLFRILSNDNDIKSRATWANAAGILKLN 488
Db 234 VFEG-----NVAGE-----RGSHTVGAAGELGPPVPGHEDVGARQVGCIG-LAVA 278
QY 489 KDEKNPKVAELYSPLISAPMVSDEIERPNVVKLGKYYLFAATRLNRGNSDDAMNANYA 548
Db 279 KOLSGEE-WEILPPLVAVGVNDQTERPHVYFODGKYIYLFITISH-----KFTYA 326
QY 549 VG-----DNVAMGVYVADSLTSGKPLNDGSLVLTASVPANWRTATYVYAVPVAGKDDQV 604
Db 327 EGTGTPDGV--YGFVGEHLFGPPYRPNASGLVL--GNPPEQFQYVSHCVMP-----NG 376
QY 605 LVTSYMTNRNGVAGKMD-----STWAPSFLLOINPNT 638
Db 377 LVTSFTI--DSVPTGEGEDYRIGGTEAPTIVRIILKGDRS 411

RESULT 11
SACB_ACEDI
ID SACB_ACEDI STANDARD; PRT; 584 AA.
AC Q43998;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-NAR-2002 (Rel. 41, Last annotation update)
DE Levanucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl
transferase) (sucrose 6-fructosyl transferase).
GN LSCA.
QS Acetobacter diazotrophicus (Gluconacetobacter diazotrophicus).
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
OC Gluconacetobacter.
OX NCBI_TaxID=33996;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 52-61.
RC STRAIN-SRT4 / CBS 550.94;
RX MEDLINE=9625399; PubMed=8704949;
RA Arrieta J., Hernandez L., Cosco A., Suarez V., Balmeri E.,
RA Menendez C., Petit-Glatron M.-F., Chambert R., Selman-Housein G.;
RT "Molecular characterization of the levanucrase gene from the
endophytic sugarcane bacterium Acetobacter diazotrophicus SRT4.";
RL Microbiology 142:1077-1085 (1996).
CC -!- FUNCTION: RELEASES FRUCTOOLIGOSACCHARIDES AND LEVAN. A HIGH-
MOLECULAR-MASS FRUCTOSYL POLYMER, FROM SUCROSE. IT ACTS MORE AS A
SUCROSE HYDROLASE THAN AS A FRUCTAN POLYMERASE.
CC -!- CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL)(N) =
GLUCOSE + (2,6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN
ALSO ACT AS FRUCTOSYL ACCEPTORS).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- PTM: THE N-TERMINAL IS BLOCKED.
CC -!- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DB EMBL; L41732; AAB36606.1;
DR InterPro; IPR003469; Glyco_hydro.68.
DR Pfam; PF02435; Glyco_hydro.68; 1.
KW Transferase; Glycosyltransferase; Signal.
FT SIGNAL 1 30 POTENTIAL.
FT PROPEP 31 51
FT CHAIN 52 584 LEVANSUCRASE.
SQ SEQUENCE 584 AA; 63525 MW; CBFBF4139AD0B8CE CRC64;

```

Query Match 6.4%; Score 266; DB 1; Length 584;  
 Best Local Similarity 26.9%; Pred. No. 1.8e-06;  
 Matches 141; Conservative 63; Mismatches 171; Indels 150; Gaps 35;

```

QY 209 TOMTY---NDFO---KIADTL-IKQDGRVYVFFKASEIKNPAATTK---DAQTNTIETP 258
Db 74 TQOAYDPQSDFTARWTRADALQIKAHSDATV-----AAGNSLPAOLTPNPADFPVYINP 129
QY 259 -LDVWDSNPQVDVRFGQVANNNGYQLYIAMGIPNQ-----NDNHIY-----LLYNKYGDNE 309
Db 130 DVMVWDWTLLIDKHADQF-SYNGWEVIFCLTADPNAGYGFDDRHVHARIGFFYRAGIPA 188
QY 310 LSH-----WKNVGPITFGYNSTAV-----SOEWSGSAVLNS--DNSIOLFTRV--- 350
Db 189 SRRPVNGWTYGGHLPDGCASAQVYAGOTYTNQAEWSSGSSRLMQIHGNTVSFVYTDVAFN 248
QY 351 -DTSNN--TNHQKTASATLYLTNNNGVSLAQVRNDY--IVFEG-----DGYVYO 396
Db 249 RDANANNITPPQAIITQT-----LGRIHADFNHVFETGTAHTPLLPDGVLYO 297
QY 397 TYDQWKATNGADN--IAMDRAHVIEDGN--GDRYLVEFAST-----GLENYQGEDOIYNW 448
Db 298 -----NGAQNEEFNFRDPTTFEDPKHPGVNVMVFEAGTAGQRGVANCTEAD----- 343
QY 449 LNYGDDAFNIKSLFRILSNDNDIKSRATWANAAGILKLNKDEKNPKVAELYSPLISAPM 508
Db 344 LGFRNDP-NAETLQEVLD-----SGAYYQKANIG-LAIATDSTLSK-WKFLSPLISANC 395
QY 509 VSDIERPNVVKLGKYYLFAATRLNRGNSDDAMNANYAVG-----DNVAMGVYVADSLT 564
Db 396 VNDQTERPQVVLHNGKYIYFTISH-----RTTFAAGYVDGPDGV--YGFVGDGIR 442
QY 565 GSKPLNDGSLVLTASVPANWRTA-----TYSYVAVPVAGKDDQVLT 607
Db 443 SDFQPMN-YGSLTGMGNTDNLTAAGTDFDPSDQNPRAFQSYSHYVMPGG-----LVE 495
QY 608 SY---MTNRNGVAGKMDSTWAPSFLLOINPNTTIVLAKMTNQ 649
Db 496 SFIDTVENRRG-----GTLAPTIVRRI-AONASAVDLRYNGG 532

RESULT 12
SACB_PSESG
ID SACB_PSESG STANDARD; PRT; 415 AA.
AC O52408;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Levanucrase (EC 2.4.1.10) (Beta-D-fructofuranosyl transferase)
GN LSC.
OS Pseudomonas syringae (pv. glycinea).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=318;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PG4180;
RX MEDLINE=98394981; PubMed=3726857;
RA Hettwer U., Jaekel F.R., Boch J., Meyer M., Rudolph K., Ullrich M.S.;
RT "Cloning, nucleotide sequence, and expression in Escherichia coli of
levanucrase genes from the plant pathogens Pseudomonas syringae pv.
glycinea and P. syringae pv. phaseolicola.";
RL Appl. Environ. Microbiol. 64:3180-3187 (1998).
CC -!- CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL)(N) =
GLUCOSE + (2,6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN
ALSO ACT AS FRUCTOSYL ACCEPTORS).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----

```



```

CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL; AF037443; AAC36056.1; -
DR InterPro; IPR003469; Glyco_hydro_68.
DR Pfam; PF02435; Glyco_hydro_68; 1.
DR KW Transferrase; Glycosyltransferase.
DR SQ SEQUENCE 415 AA; 45844 MW; 839B686AC80610CF CRC64;

Query Match
Best Local Similarity 6.2%; Score 257; DB 1; Length 415;
Matches 103; Conservative 66; Mismatches 152; Indels 88; Gaps 19;

QY 261 VMDSWPQDVRTGOVANNNGYOLVIAVMGIPNQND-----NHVLL-----YNKYCDNEL 310
DB 44 IWTMPLELREL-DGTVSVNGSVIVTLTADRHDPDPQYVGANGRVDIRKDWEDRGRARM 102
QY 311 SHWKN-----VGPIFGYNSTAVSOEWSSAV-LNSDNGSTQLFTRVDTSDNNTNHO 361
DB 103 CYWYSRTGKDWIFGGRVWAGVSPTRWAGTPVLLNDKGDIDLYTCV-----TPGAA 156
QY 362 IASATLYLTDNNGVSLAQVRNDYIVFEGDGYVQYQYQWKATNGKADNIARD-AHVIE 420
DB 157 IAKVRGRIVTSKGVKELADFTVKTLPFADKGYQYQTEAQNSTW-----FRDPSPFID 209
QY 421 DCGDRYLVFEASTGLENYQGEDQIYNWLNNGGDAFNKLSFRILSNDDIKSRATWANA 480
DB 210 PNDGKLYWVFE-----NVAGE-----RGTHVGAELGPVPFGHEETGARGFVG 255
QY 481 AIGIL---KLNKDEKPKVAPLYSLISAPMVSEIERPNVYKNGKYYLFAATRLNRGS 537
DB 256 CIGLAVAKDLSDGE-----WEILLPLVAVGVNDQTERPHYVFDGKYLYFT----- 302
QY 538 NDDAWMANAVAGDNV-----AMGVYVADSLTSGYKPLNDGGVLTASVPANWRTATYSY 593
DB 303 -----ISHKTYADGVTPGDPGVYGFVGEHLFGPYRPNASSGLV-LGNPPAQ-PFQIYSHC 355
QY 594 AVPVAGKDDQVLTYSYNNRNGVAGKMD-----STWAPSLLQINDPNT 638
DB 356 VMP-----NGLVTSFI---DSVPTSGDYRIGGTGTEAPTVRILLEGDRS 395

RESULT 13
FNBA_STAAU
AC FNBA_STAAU : STANDARD; PRT: 1018 AA.
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Fibronectin-binding protein precursor (FNBP).
GN FNBA.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIC 8325-4;
RX MEDLINE=89098998; PubMed=2521391;
RA Signaes C., Raucii G., Joensson K., Lindgren P.-E.,
RA Anantharamiah G.M., Hoeck M., Lindberg M.;
RT "Nucleotide sequence of the gene for a fibronectin-binding protein
RT from Staphylococcus aureus: use of this peptide sequence in the
RL synthesis of biologically active peptides."
RL Proc. Natl. Acad. Sci. U.S.A. 86:699-703(1989).
CC -1- FUNCTION: THE ABILITY OF BACTERIA TO BIND FIBRONECTIN HAS BEEN
CC PROPOSED AS A VIRULENCE FACTOR ENABLING BACTERIA TO COLONIZE
CC WOUND TISSUES AND BLOOD CLOTS. BINDING OF PLASMA FIBRONECTIN TO
CC THE BACTERIAL SURFACE MIGHT BLOCK ADHESION RECEPTORS ON S.AUREUS,
CC THUS REPRESENTING AN IMPORTANT DEFENSE MECHANISM AGAINST TISSUE
CC INVASION.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS

```

```

CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL; J04151; AAA26632.1; -
DR InterPro; IPR004237; Fn_bind.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF02986; Fn_bind; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
DR SIGNAL; Repeat; Cell wall; Transmembrane.
FT CHAIN 1 36
FT DOMAIN 37 1018 FIBRONECTIN-BINDING PROTEIN.
FT TRANSMEM 994 1009 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 1010 1018 MEMBRANE ANCHOR.
FT REPEAT 545 574 CYTOPLASMIC (POTENTIAL).
FT REPEAT 575 604 B-1.
FT DOMAIN 745 878 B-2.
FT REPEAT 745 878 4 X APPROXIMATE TANDEM REPEATS,
FT REPEAT 783 820 FIBRONECTIN-BINDING DOMAIN.
FT REPEAT 821 859 D-1.
FT REPEAT 860 878 D-2.
FT DOMAIN 879 948 D-4 (INCOMPLETE).
FT REPEAT 893 906 5 X TANDEM REPEATS, PRO-RICH (WR).
FT REPEAT 907 920 WR1.
FT REPEAT 921 934 WR2.
FT REPEAT 935 948 WR3.
FT DOMAIN 962 987 WR4.
FT REPEAT 987 987 WR5.
FT CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT PROTEINS.
SQ SEQUENCE 1018 AA; 111780 MW; 58175E0020E81F1F CRC64;

Query Match
Best Local Similarity 6.2%; Score 255; DB 1; Length 1018;
Matches 195; Conservative 146; Mismatches 322; Indels 398; Gaps 50;

QY 55 VTLSGGQVSAAD---TTIRTSANANSASSAANTONSQV-----ASSAITSSTS 104
DB 27 VGMGQDKEAAASEQKTTVEENGHSATDKTSETQTATTWYNHIEETQSYNATVTEQPSN 86
QY 105 AASLNNTDSKAAQENTWTAK--NDDTKAAPANESSEAKNEPAVNVNDSSAANKDDQSS 162
DB 87 ATQVTEEAPKAVQAPQTAQAPANIETVKEEVKEEAPQVKEITQSDQSDGQDQVLDLP 146
QY 163 KNTTAKLNKDAENYVKAGIDPNSLTDDOIKALNKWNFSKA-----AKSGTMT--- 212
DB 147 KKATQ---NQVAETQVEVA--QPTASESKPRVTRSDVAEAKESNAKVEGTDTVTSKV 201
QY 213 -----YDFQKIAD-----TLIK-----QDGRYTVPEF-----K 236
DB 202 TVEIGSIEGHNNNTNKVEPHAGRAVLKYLKPEGLHQGDYDFTLSSNNVNHGYSTARK 261
QY 237 ASBKN-----MPATTKDAQTWIE--PLDVMDSWPVQ---DVRTGOVANNNGVQ 282
DB 262 VPEIKNGSVMATGEVLEGKRYRTFTNDIEDKVDVDAEINLFDPKTQV---NGNQ 318
QY 283 LVIAAM-----GIPN--QNDNHYLLYKNGDNLSHKNGVPIFGYNS-- 324
DB 319 TITSTINEQTSKELDVYKDGIGYVYANLAGSIETFNK--ANNRESHVAFIKPNNGKTS 377
QY 325 -TAVSQWSSGSAVLNSDNIQLFYTRVDTSD-----NNTNHQIKASATLYL- ---D 371
DB 378 VTVTTLKNGNQNGPKVRIFEYLGNNEDIKSVYANTTDTSKFEVTSNMSGNLQ 437

```

```

QY 372 NNGMNSL--AQRNDYIVFEGDGYQYQYDQWKATNKAGDNAMRDARHVIEDGNGDRYL 429
Db 438 NNGSYSLNENLDKTYVY-HDGEYL-----NCTDEVDFRTQMV--GH----- 477
QY 430 FEASTGLENYGEDQIYNLNYGGDAFNKSLFRILSNDDIKSRATWANAII----- 482
Db 478 -----PEOLY---KYVYDRGYTL-----TWDNGLVLYSKNAN 506
QY 483 -----GILKNKDB-KNPKVAELYSPLISAPWYSDIERPNVVKLGKNYLFAATRLNR 535
Db 507 GNEKNQPIONKEFEYEDTIKETLQYDKNLVTTVEEYDSSTLIDYH-----TAIDG 562
QY 536 G-----SNDAMWNNANT--AVGDNVAVWG----- 557
Db 563 GGYVDGYIETIETDSADIDYHTAVDSEAGHVGGYTESSESNPIDFEESTHENSXH 622
QY 558 -----YVADSLGYSKPLNDGCV-----LTASVPANWRATATSYAV----- 595
Db 623 HADVVEEEDTNPGGGVTTESNLVFEDEESTKGIVTGAVSDHTTVEDTKYTTESNLIE 682
QY 596 -----PVAGKDDQVLVTSYMTNRNGVAGKGMDSWAPSFLLQINPDNTTTLVAKM 645
Db 683 LYDELPEHQAGQPVVEETIKNNHLSHSLGTENGHNY--DVIEEIEENSHVDIKSEL 740
QY 646 -----TNQGDWIDDSSENLDM-----IGDLSAALP-----GERDK 677
Db 741 GYEGGQNGSGNOSFEETEE-DKPKYBQGGNIYDIDFDSVPQIHGQNGKNGSFEEDTKDK 799
QY 678 P-----VDMDLIG--YGLKPH----- 691
Db 800 PKYEHGNIIDFDSVPHIGHFNKHTIEEDTNKDKPSYQFGHNSVDFEEDLPKVS 859
QY 692 -----DPATNDPPTPT--TPETPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 726
Db 860 GQNEGQOTTEEDTTPVPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 918
QY 727 QTPNTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 750
Db 919 ETP-TPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 750
QY 751 TNNRLPOTG--NNANKAMIGLGMGTLLSMFGLAEINKRRFN 789
Db 978 KASLEPTEGGESTNKGML-----FGGLFSILGALLRRNKN 1015

RESULT 14
BAG_STRAG
ID BAG_STRAG STANDARD; PRT; 1164 AA.
AC P27951;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Iga FC receptor precursor (Beta antigen) (B antigen).
GN BAG.
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1311;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-48.
RC STRAIN=LA239;
RX MEDLINE=91312121; PubMed=1857207;
RA Jerlstrom P.G., Chhatwal G.S., Timmis K.N.;
RT "The Iga-binding beta antigen of the c protein complex of Group B
RT streptococci: sequence determination of its gene and detection of two
RL binding regions.";
RL Mol. Microbiol. 5:843-849(1991).
RN [2]
RP IDENTIFICATION OF IG-LIKE DOMAIN.
RX MEDLINE=97035265; PubMed=8880921;
RA Bateman A., Eddy S.R., Chochia C.;
RT "Members of the immunoglobulin superfamily in bacteria.";
RL Protein Sci. 5:1939-1942(1996).

```

```

CC -|- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -|- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE DOMAIN.
CC -|- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X59771; CAA42442.1;
DR PIR: S15330; FCSOAG.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00746; Gram_pos_1;
DR SMART: SM00409; IG_1;
DR PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
KW Cell wall; Transmembrane; Receptor; Repeat; Signal;
KW Immunoglobulin domain.
FT SIGNAL 1 37
FT CHAIN 38 1164 IGA FC RECEPTOR.
FT DOMAIN 38 1131 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1141 1159 MEMBRANE ANCHOR (POTENTIAL).
FT DOMAIN 1160 1164 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 434 534 IG-LIKE DOMAIN.
FT DOMAIN 199 438 IGA-BINDING (POTENTIAL).
FT DOMAIN 439 826 IGA-BINDING (POTENTIAL).
FT DOMAIN 827 945 PRO-RICH REPEATS.
FT DOMAIN 1131 1137 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT SEQUENCE 1164 AA; 131051 MW; 65DE94AF720A5474 CRC64;
SQ

```

```

Query Match 5.8%; Score 241; DB 1; Length 1164;
Best Local Similarity 20.0%; Pred. No. 7.8e-05;
Matches 191; Conservative 130; Mismatches 347; Indels 288; Gaps 40;

QY 10 VVTLSAALVFGA-----TTVNAS-----ADTNINNDSSST-----VQV 43
Db 21 VASVAVASLFGMSVAHASLVKDDSVKTEVAAKPPSMAOTDQGNSSSSSELETTKMEI 80
QY 44 TTCDNDIAKSVTLGSGQVSAASDTTIRTS-----NANSASSAANTONSNSQVASSAAT 99
Db 81 PTTDIKKAPEVPEKTAGETS-ATDTGKREKOLQOWKNLKNVDVNTILSHQKNEFKTKI 139
QY 100 SSTSSAASL-----NNT-----DSKAAQENTNTAKNDTO----- 129
Db 140 DETNDSDALLELQNFNETNRLHLHKQHEEVEKDKKAKQ--KTLKOSDTKVDLSNIDKE 197
QY 130 ---KAAPANESEAK---NEPAVNVNDSAAKND-----QOSSKK----- 164
Db 138 LNHQKSQVEMAEQKGTINED-----KDSMLKKIEDIRKQAQADKDEAEVKEELGKL 253
QY 165 ---NTTAKLNKDAENVVK-----AGID---PNSLTDQIKALNKMFSAKSGTOM 211
Db 254 FSSTKAGLDQIOEHVVKETSEENTOKVDEHYANSLQAQKSLSELDKATTNEQATQV 313
QY 212 TYNDF---OKIAD--TLIKODGRYTVPFKA-----SEIKNMPATTKDAQTNTI 256
Db 314 K-NQFLENAOKLEIQPLIKET---NVKLYKAMSESLQEVEKELKHNSLEANLADVAKSK 369
QY 257 EPLDWDVS----- 370
Db 370 EIVREYEGKLNOSKNLPKOLKEEEAHSKLQYVEDFRKFKTSEQVTPKRVRDLAAN 429
QY 278 WNGYQ---LVIAHMGIPNQNNDNHYLLYNNKYGDNELSHWKNVGPFGYNSTAVSOWSGS 334
Db 430 ENNOOKIELTVSPENITVVEGEDVKETVAKSDSKIT--LDFSDDLTKYNPSVSDRISTN 487
QY 335 AVLNSDNSITOLFTRVDTSDNNTNHQIKASATLYLTDNNGNVSLAQVRNDYIVFEGDGY 394

```

Db	488	YKNTON	----	HKIAETIKNLKNESQT	VT	LKAKD	SGNV	----	VEKFT	ITV	----	533
QY	395	YOTYQW	KATNKAGD	NIAMRAH	VI	----	----	----	----	----	----	445
Db	534	----	QKKEEK	QVPTPKQ	OKS	TEEK	VPQ	PKPSND	KNO	LQELIK	KAQ	583
QY	446	YNW	LYNGD	DAFN	IKSLFR	ILS	ND	DIKS	----	RAT	WAN	491
Db	584	----	----	----	----	----	----	----	----	----	----	584
QY	492	KNP	KAELYS	PLISAP	WVS	DS	IE	RP	----	----	----	629
Db	630	SKRY	TEHYFN	KYS	DF	MY	LO	HAQ	ME	MLTR	KV	645
QY	546	NY	AVG	DN	AM	GY	VAD	SL	TG	SY	K	601
Db	690	NY	GLS	END	AL	KGY	E	KY	E	F	LP	739
QY	602	DQ	----	----	----	----	----	----	----	----	----	655
Db	740	DO	AK	E	K	A	I	A	V	S	K	780
QY	656	SSEN	LD	M	G	L	D	S	A	A	L	702
Db	781	E	A	I	K	O	T	T	F	D	I	839
QY	703	T	T	P	E	T	P	E	T	P	E	758
Db	840	Q	A	P	D	T	P	Q	A	P	D	895

RESULT 15

SSP5\_STRGN STANDARD; PRT; 1500 AA.

ID SSP5\_STRGN STANDARD; PRT; 1500 AA.

AC P16952; 054184;

DT 01-AUG-1990 (Rel. 15, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Agglutinin receptor precursor (SSP-5).

GN SSP5 OR SSPB.

OS Streptococcus gordonii.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1302;

RI [1]

RP SEQUENCE FROM N.A.

RC STRAIN=M5;

RX MEDLINE=90236997; PubMed=2185241;

RA Demuth D.R., Golub E.E., Malamud D.;

RT "Streptococcal-host interactions. Structural and functional analysis

RT of a Streptococcus sanguis receptor for a human salivary

RT glycoprotein.";

RL J. Biol. Chem. 265:7120-7126(1990).

RI [2]

RP REVISIONS, SEQUENCE FROM N.A.

RC STRAIN=M5;

RX MEDLINE=96310377; PubMed=6733238;

RA Demuth D.R., Duan Y., Brooks W., Holmes A.R., McNab R.,

RA Jenkins H.F.;

RT "Tandem genes encode cell-surface polypeptides SspA and SspB which

RT mediate adhesion of the oral bacterium Streptococcus gordonii to

RT human and bacterial receptors.";

RL Mol. Microbiol. 20:403-413(1996).

CC -1- FUNCTION: MAY BIND SIALIC ACID RESIDUES OF SALIVARY AGGLUTININ

CC (SAG) IN A CALCIUM-DEPENDENT REACTION. THE INTERACTION OF SAG WITH

CC ITS RECEPTOR IN VARIOUS ORAL STREPTOCOCCI MODULATE BACTERIAL

CC COLONIZATION OF ORAL TISSUE AND IS ASSOCIATED WITH REDUCED LEVELS

CC OF DENTAL CARIES.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.

CC -1- DOMAIN: THE PR2 REGION, BY SIMILARITY WITH THE PROLINE RICH

CC DOMAINS OF THE S.PHAGES M6 PROTEIN AND STAPHYLOCOCCAL PROTEIN A,

CC MAY TRAVERSE THE CELL WALL PEPTIDOGLYCAN AND IS FOLLOWED BY

HYDROPHOBIC AA WHICH MAY FUNCTION TO ANCHOR THE PROTEIN TO THE MEMBRANE.

-1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS IN THE REGION OF THE MEMBRANE ANCHOR.

-1- SIMILARITY: BELONGS TO THE SPAP/SPSP5/SPAA FAMILY.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; U40026; AAC44100.1; -

PIR; A35186; A35186.

InterPro; IPR001899; Gram\_pos\_anchor.

Prfam; PF00746; Gram\_pos\_anchor; 1.

PROSITE; PS00343; GRAM\_POS\_ANCHORING; 1.

Signal; Repeat; Calcium-binding; Transmembrane.

SIGNAL 1 38

FT CHAIN 39 1500

FT DOMAIN 39 1474

FT TRANSMEM 1475 1495

FT DOMAIN 1496 1500

FT SIMILAR 164 470

FT DOMAIN 164 470

FT REPEAT 164 241

FT REPEAT 242 323

FT REPEAT 324 405

FT REPEAT 406 470

FT DOMAIN 771 887

FT DOMAIN 1414 1436

FT DOMAIN 1467 1472

CA\_BIND 220 235

CA\_BIND 301 316

CA\_BIND 931 950

CA\_BIND 1300 1315

SEQUENCE 1500 AA, 164552 MW; DCF190E7D44D889F CRC64;

AGGLUTININ RECEPTOR.

EXTRACELLULAR (POTENTIAL).

POTENTIAL.

CYTOPLASMIC (POTENTIAL).

TO M PROTEIN OF S.PYROGENES.

4 X APPROXIMATE TANDEM REPEATS, HRL.

1.

2.

3.

3 X APPROXIMATE TANDEM REPEATS, PRI.

PRO-RICH (PR2).

CONSERVED IN GRAM-POSITIVE COCCI SURFACE PROTEINS.

POTENTIAL.

POTENTIAL.

POTENTIAL.

POTENTIAL.

```

Query Match      5.6%; Score 229.5; DB 1; Length 1500;
Best Local Similarity 20.7%; Pred. No. 0.00041;
Matches 196; Conservative 120; Mismatches 368; Indels 265; Gaps 45;

QY      3 KSGKNWAVVLTSLAAVFGATTYNASADTNIENNDSSTVQVTTGDNNDIAVKSVTLSGGV 62
Db      14 KVAKTLCGAVLGTALAFADKAV--FADEVTFETSTSTVEVATTGN--PATNLPEAQGEM 69

QY      63 SAASDTTTIRTSANASASSANTQNSQVASSAAITSTSTSSAASLNNT--DSKAQENTN 121
Db      70 SOYAK---ESQAKAGSKESALPVEVSSADLDKAVADAKSAGYKVQDETKDKGTATTATD 126

QY      122 TAKNDTQKAAAPANESSEA-----KNEPAVN-----VNDSSAAKNDQOSSKK----- 164
Db      127 NAGQDEIKSDYAKQAEIKTTTEAYKVEVAHQAHOAETDKINAENKAADKYQKDLKSHOE 186

QY      165 -----NTTAKLNKDAE-----NVVKRAGIDPN-----SLTDQQLAKLNKN-- 200
Db      187 EVEKINTANATAKAEYEAQLAQYQKDLATVKKANEDSQOQYONKLSAYQTELARVQKANA 246

QY      201 -----FSKAKSGTOMTYNDFOKIADTLTKDGRVTVPFFKASEIKNMPAATTKDAQNT 255
Db      247 EAKEAYEKAVKENT--AKNEALKVENEALIKQ-----RNETAKATVEAAKKQ 290

QY      256 IPELDVWDSWPQDVRTGOVANWN---GYQLVIAMGIPNODNHNHYLLYNK----- 304
Db      291 YEA-----DLAAIKKANEDNDADYQAKLAAYQTELARVQKANAEEADKAVKENTAKNT 346

QY      305 --YGDNELSHWKNVGP IFGYNSTAVSQEWSGAVLNSDNSQLFTVRVDTSDNNNTNHQKI 362
Db      347 AQAENEALKQRNETAKATYDAAVKKYEAADLAQVKQAN-----ATNEADYV-- 392

```

QY	363	ASATFLYLTDDNNGNVSLAORNDYIVPEGDGYTYOTYDOWKATNKGADNIAMRDHAHVEDS	422
Db	393	AKLAAQT-----LARVO-----KANADAKATYKAVEDNKKAKNAALKAENEIEIKOR	440
QY	423	NGDRYLVFFEA-----STGLENYOGEDQTY-----NWJLNYGGDDAFNIKSLF	463
Db	441	NAVAKTDYERAKLAKYEAZDLAKYKKEFAAYTAALAEAESKKQDGYLSEPRSOSLNFKSEP	500
QY	464	RILSND-----IKSRATWANAATGILKLANKDENKPKVAELISPLIS-----	505
Db	501	NAIRTIJDSVHOYGOQELDALVKS-----W-----GISPTNDRKKSRAYSIFYNAINSNNTY	552
QY	506	APMWDEIERNNVYKLGKYYLPAFTLNR-----GSND-----DA	541
Db	553	AKLVLEKDKPDVITYTGLKNSFNGKRIKSVVITYTKETGFENDTKMTMFASSDPTVTA	612
QY	542	WNNANYAVGDNVAMVGVAD-----SLAG-----SYKPLNDSQSVLITASVPANWRTAT	589
Db	613	WYN-DYFTSTINNVKRYFDEEQGLMNLTLGVLNFGSLNRGSGAIDKDAIES-----	668
QY	590	YSYAVPVAGKDDQVLV-----TSYMTNRNGVAGKMDSTWAPS-FILQINPDNTTIVLAKMT	646
Db	669	FNGRYIPIISGSSIKIHENNSAYADSSN-AEKSLGARNTSEWDTTSSPNWYGAIVGEI	726
QY	647	NOGDW-----IWDSSENLDMTGDDLSAALPG-----ERDKPVDMDLI--GY	686
Db	727	TOSEISFNMASSKSGHINWAFNSINAIGVPTKPVAPTQPMYETEKELEPAPVAPSY	786
QY	687	GLKPHDPA-TPNDPEPTPTTP-----ETNTP--KTPKTPENPGTTP	726
Db	787	ENETPPVKTPOQE-PSKPEPTVETEKELEPAPVAPSYENETPPVKTIPOQE-PSKP	844
QY	727	QTP-----NTNTP-EIPLPETPKQETOTNNRLP	756
Db	845	EEPNETEKELEPAPVAPSYENETPPVKTIPOQE-PSKEEPTYDPLP	892

Search completed: September 26, 2002, 18:26:43  
Job time: 182 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2002, 18:23:41 ; Search time 40.2 seconds  
(without alignments)  
1885.932 Million cell updates/sec

Title: US-09-995-587A-1

Perfect score: 4129  
Sequence: 1 MYSGKNWAVVTLSTAAALVF.....GMCTLLSMFGLAEINKRRFN 789

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR.71.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2016.5	48.8	797	2 B28551	levansucrase (EC 2
2	744.5	18.0	473	2 A25040	levansucrase (EC 2
3	741.5	18.0	472	2 J00802	levansucrase (EC 2
4	707	17.1	489	2 F97118	levansucrase [impo
5	356.5	8.6	428	2 H97118	levansucrase [impo
6	299	7.2	423	2 S33771	levansucrase - Zym
7	298.5	7.2	413	2 JC2520	beta-fructofuranos
8	297	7.2	423	2 JC2519	levansucrase (EC 2
9	291.5	7.1	413	2 S47527	extracellular suc
10	281	6.8	415	2 S39195	levansucrase - Erw
11	261	6.3	1038	2 H90053	hypothetical prote
12	255	6.2	1018	2 A32192	fibronectin-bindin
13	246	6.0	940	2 S19702	IgA Fc receptor pr
14	241	5.8	1164	1 FCSOAG	hypothetical prote
15	237.5	5.8	1134	2 A60234	hypothetical prote
16	231.5	5.6	961	2 B90053	hypothetical prote
17	231.5	5.6	2551	2 B98047	salivary agglutini
18	230	5.6	1661	2 A35186	dextranase - Strep
19	229.5	5.6	1473	2 A35186	probable peptidogl
20	227	5.5	1337	2 T30291	protein F4N2.10 [i
21	225	5.4	2044	2 AB1180	hypothetical prote
22	224	5.4	138	2 D96715	sporozoite surface
23	224	5.4	1072	2 A86827	FmcB protein [impo
24	220	5.3	826	2 A45559	saliva-interacting
25	219	5.3	2481	2 D90011	uncharacterized pr
26	215	5.2	1556	2 A60988	Ig light chain-bin
27	213	5.2	1565	2 S04729	
28	212	5.1	2817	2 B97033	
29	211	5.1	719	2 A42808	

30 210.5 5.1 1166 2 T28680  
31 208.5 5.0 1256 1 A43829  
32 208.5 5.0 1463 2 T30290  
33 206 5.0 1092 2 T30214  
34 205 5.0 1315 2 T28679  
35 202.5 4.9 430 2 JC3739  
36 202 4.9 1449 2 T30552  
37 199 4.8 1093 2 B86748  
38 199 4.8 1592 2 A38175  
39 198.5 4.8 357 2 S21758  
40 197 4.8 877 2 F90070  
41 196.5 4.8 1025 2 S69790  
42 196 4.7 792 2 S70305  
43 195.5 4.7 1193 2 B96943  
44 195.5 4.7 1449 2 T30857  
45 195 4.7 879 2 S23006

## ALIGNMENTS

RESULT 1

B28551

levansucrase (EC 2.4.1.10) precursor - Streptococcus mutans (strain GS-5)  
N:Alternate names: sucrose 6-fructosyl transferase  
C:Species: Streptococcus mutans  
C>Date: 08-Mar-1989 #sequence\_revision 08-Mar-1989 #text\_change 15-Oct-1999  
C:Accession: B28551  
R:Shiroza, T.; Kuramitsu, H.K.  
J. Bacteriol. 170, 810-816, 1988  
A:Title: Sequence analysis of the Streptococcus mutans fructosyltransferase gene an  
A:Reference number: A91892; MUID:88115184  
A:Accession: B28551  
A:Molecule type: DNA  
A:Residues: 1-797 <SHI>  
A:Cross-references: GB:M18954; NID:g153635; PIDN:AAA88584.1; PID:g153636  
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 48.8%; Score 2016.5; DB 2; Length 797;  
Best Local Similarity 50.4%; Pred. No. 9.1e-97;  
Matches 408; Conservative 118; Mismatches 211; Indels 73; Gaps 14;

QY 1 MYSGKNWAVVTLSTAAALV-FGATVTVNASADTNIEENDSSTVQVTTGNDIAVRSVTILGS 59  
DB 9 MYKKGFWVAVITTTAMLTGIGLSSVQA-----DEANSTQV-----S 45  
QY 60 GOVSAASDTTITSANASSAANTQNSOVASSAATSTSSAASLNNTD--SKAAQ 117  
DB 46 SELAERSQVQENTTSSSSAAENQAETVQET-----PSTNPAATVNTDQTKVIT 97  
QY 118 ENTNTAKNDTQKAPANESSEAKNEPAV---NVNDSAAKNDQOOSKKNTTAK--LNK 172  
DB 98 DNAAVESKASKTKDQAATVTTAASTPEVGQTNKAKATKADITTPKNTIDEYGLTE 157  
QY 173 DAENVVYKAGIDPNSLTDQKALNKNMNFSAKASGTMQTYNDFQKIADTLIKODGRYTV 232  
DB 158 QARKIATEFAGINLSLTQKQVEALNKKVLTSDAQTGHQMTYQEFDKIAQTLLAQDERYAI 217  
QY 233 PFFKASETKNNPAATTKDAQNTTIEPLDWDSPYQDVYRGQVAVNNGYQLVIAAMGIPN 292  
DB 218 PYFNAKAIKNKAATTKDAQTGQIADLDVWDSPVQAKTGEVINWNGYQLVAVAMGIPN 277  
QY 293 QNDNHIYLLNKKYGDNLSLHKNVGPFGYNSTAVSOWSGSAVLNSDSNLSIQLFYTRVDFT 352  
DB 278 TNDNHIYLLNKKYGDNDFHKNAGSIFGYNELPTQWESGSATVNEGGSLQLFYTKVDFT 337  
QY 353 SDNNTNOKTASATLYLTNDNNGVSLAQVRNDIVF-EG-DGYYYQTYDQWKATNKGAON 410  
DB 338 SDKSNNRQATATVNYGFGDDQVRLSVENDKVLTPEGVNAVHYQSYQQRSTFTGADN 397  
QY 411 IAHRAHVIEDGNGDRYLFPEASTGQENYQGEQIYNWLNKGGDDAFNFKSLFRILSNDD 470

S. T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.

## C:Genetics:

A:Gene: sacB  
 C:Keywords: glycosyltransferase, hexosyltransferase  
 F:1-29/Domain: signal sequence #status predicted <SIG>  
 F:30-472/Product: levansucrase #status predicted <MAT>

Query Match 18.0%; Score 741.5; DB 2; Length 472;  
 Best Local Similarity 38.1%; Pred. No. 3.5e-31;  
 Matches 188; Conservative 85; Mismatches 172; Indels 49; Gaps 18;

QY 175 ENVVKAGI---DPSNLDDQIKALKNMNFSSKAASG---TQMTYNDFOKIADTLIKODG 228  
 DB 4 KKVROATVLTFTTALLAGGATQAFKENNQKAYKETGVSHITRDMLOPKQ--QQNE 61  
 QY 229 RYTVFFKASEIKNMPAATTKDAQNTTIEPLDVDSWPDVQVGVQVGNWNGYOLVIAMM 288  
 DB 62 KYQVPOFQOSTIKNIESA-----KGLDVDSWPLQNA-DGTVAEYNGYHVVFALA 110  
 QY 289 GIP-NQNDNHLYLLNRYKGDNELSHWKNVGP-----FGYNSTAV---SOEWSGSVALN 338  
 DB 111 GSPKADDDTSYMFYQYKVDNSIDSWKNAGRVFKDSKFDANDPILKDOTQEWSESGATFT 170  
 QY 339 SDNSTOLFTYTRVSDNNTNHHQKTASATLYLTNDNGVNSLAQVRNDYIVFEGDGYYYQTY 398  
 DB 171 SDGKIRLFT--DYSGRHYGKQSLTTAQVNVYSKSDDTLKNVGDHKTIFDGDGKTYQNV 228  
 QY 399 DQW--KATNKAGADNIAHRDVIEDGNGDRYLVFEASTGLN-YQGEDQIYNLNLNGYGGD 455  
 DB 229 QQFIDEGNTSGDNHTRDPHYVED-KGKYLVEANTGTENGYYQGESLFNKAYYGGGT 287  
 QY 456 AFNKSIFRLSNDNDIKSRATWANAAGILKLNKDEKNPKVAELYSPLISAPMVSDIEIR 515  
 DB 288 NFFKESQK-LOOSAKKRDAAELANGALGILNNDYTLKKVMK---PLITSNTVTDEIER 343  
 QY 516 PNWKLGNKYLLFAATRLNRSGNDAMNANAVGDNVAVGVSADSLTSGYKPLNDSGV 575  
 DB 344 ANVEKMGKWLFTDSRGSKWTIDG--INSN-----DIYMLGYVNSLTPGPKPLNKTGL 396  
 QY 576 VLTSAPVAPWRTATYIYAVPVAGKDDQVLVTSYMTNRRNGVAGKGMDSWAPSFLQLINP 635  
 DB 397 VLQGLDLPNDVTYSHFAVPOA-KGNVVITSYMTNRRNGFFEDK--KATFGSFLMNIKG 453  
 QY 636 DNTTTLAKNTNQG 649  
 DB 454 NKTSVVKNSILPQG 467

## RESULT 4

F97118  
 levansucrase [imported] - Clostridium acetobutylicum  
 C:Species: Clostridium acetobutylicum  
 C>Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
 C:Accession: F97118  
 R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee.  
 J. Bacteriol. 183, 4823-4838, 2001  
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld  
 A:Reference number: A96900; MUID:21359325; PMID:21359325  
 A:Accession: F97118  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-489 <KUR>  
 A:Cross-references: GB:AE001437; PIDN:AAK79737.1; PID:g15024742; GSPDB:GN00168  
 A:Experimental source: Clostridium acetobutylicum ATCC824  
 C:Genetics:  
 A:Gene: CAC1772

Query Match 17.1%; Score 707; DB 2; Length 489;  
 Best Local Similarity 37.2%; Pred. No. 2.2e-29;  
 Matches 183; Conservative 88; Mismatches 173; Indels 48; Gaps 20;

QY 188 LTDDQIKALKNMNFSSKAASGQMTYNDFOKIADTLIKODGRTYVFFKASEIKNMPAAT 247  
 DB 21 ITQOTFASTNDMNY-KETYGSHITRYNMSKI--PMEQNDLKFKVPQPNASTLKNIASAK 77  
 QY 248 TKDAQNTTIEPLDVDSWPDVQVGVQVGNWNGYOLVIAMGIP-NQNDNHLYLLNRYK 306  
 DB 78 GYDKNGNLID-LDVDSWPLQN-GDGTVANYHGHIIFALAGDPKNOODTSIYMFYQKIG 135  
 QY 307 DNELSHWKNVGPFGYNSTAV-----SOEWSGSVALNSDINSIOLFTRVD--TSDN 355  
 DB 136 ENSIDSWKNAGRVFKDSKYVANDPYLKYQTEWSGSATLTSQGVRLFTDFSGVAKDG 195  
 QY 356 NT--NHOKIASATLYLTNDNGN-VSLAQVRNDYIVFE-GDGYYYQTYDQWKAATNK--GAD 409  
 DB 196 GTDASNOVITTVQNLSDPSNTINSDVSHKSVFGGNGGTIYQNYQOFIDEGKWSGD 255  
 QY 410 NIAMRDVIEDGNGDRYLVFEASTGL-ENVQEDQIYNLNLNGYGGDGAENIKSLFRLSN 468  
 DB 256 NHTLRDPHYVED-NGRKYLVFEANTGTNDGYQDTSLLNKAFFGRSOSFFKTEKQDLID 314  
 QY 469 DDIKSRATWANAAGILKLNKDEKNPKVAELYSPLISAPMVSDIEIRPNVYKLGNYLYF 528  
 DB 315 TNKHKDASLANGALGILNNDYTLKKEMK---PLIASNTVTDEIERANVFNKNGRWYLF 371  
 QY 529 AATRLNRSGNDAMNANAVGDNVAVGVSADSLTSGYKPLNDSGVLTASVAPWRTA 588  
 DB 372 TD---SRGSK---MTINGISSKDIYMLGFSSNSLTGPYKPLNKTGLVNLNLDPLDTF 424  
 QY 589 TSYIYAVP-VAGKDDQVLVTSYMTNRRNGVAGKGM--DSTWAPSFLQLINPNTTTLA 643  
 DB 425 TYSHFAPQTNKGN--VWITSYITNR-----GMSDHSHPAPSFLNKAIGTKTSVISN 476  
 QY 644 KMTNQGDWTD 655  
 DB 477 SILQOQLTIDN 488

## RESULT 5

F97118  
 levansucrase [imported] - Clostridium acetobutylicum  
 C:Species: Clostridium acetobutylicum  
 C>Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
 C:Accession: F97118  
 R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.;  
 J. Bacteriol. 183, 4823-4838, 2001  
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacteriu  
 A:Reference number: A96900; MUID:21359325; PMID:21359325  
 A:Accession: F97118  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-428 <KUR>  
 A:Cross-references: GB:AE001437; PIDN:AAK79739.1; PID:g15024744; GSPDB:GN00168  
 A:Experimental source: Clostridium acetobutylicum ATCC824  
 C:Genetics:  
 A:Gene: CAC1774

Query Match 8.6%; Score 356.5; DB 2; Length 428;  
 Best Local Similarity 27.2%; Pred. No. 2.4e-11;  
 Matches 125; Conservative 70; Mismatches 176; Indels 89; Gaps 22;

QY 209 TQMTYNDFOKIADTLIKODGRTYVFF-----FKASEIKNMPAATTKDAQTNTI 256  
 DB 3 TRTY-----KMSLVVILAILTIPFLILRHNTGYTISRSQAQNFKC--TKENTAPNI 56  
 QY 257 EP-----LDVDSWPDVQVGVQVGNWNGYOLVIAMGIPN-QNDNHLYLLNRYK 306  
 DB 57 NPNFKLTAPNLWWDWTPLVK-KDGLAVVNGYKVFALTASRNWGNKRHDVAGSYFC 115  
 QY 307 DNELSHWKNVGPFGYNSTAVSOWSGSVALNSDINSIOLFTRVDSN--FTNHQKIAS 364  
 DB 116 STDGNNVYKGLAYNVEDALGSRWAGSAILDENGWVQFFYTATGRKEAVRTFEQRLVK 175



A: Type: DNA  
A: Residues: 1-423 <SN>  
A: Cross-references: GB:L08093; NID:g5880348; PIDN:AAA27695.1; PID:g295699

Query Match 7.2% Score 299; DB 2; Length 423;  
Best Local Similarity 26.68; Pred. No. 2.3e-08;  
Matches 120; Conservative 73; Mismatches 156; Indels 102; Gaps 23;

QY	250	DAQNTTIEPLD-----YWDSMPVQDVRTGQVANNGYQLVIAMM-----CIPN	292
DB	24	DQPTATMTTIDYDPVMTDKYVWDWTWPLRDI-NGVVVFGQSIVFALVADRTKYGWHN	82
QY	293	QNND-HIYLKLYKGDNELSHWRNVGPIFYCNSTAVSQEWSGSAVL--NSDNSIQLFYTR	349
DB	83	RNDGARICYPSRGGSN-----WIFGGHLLKDGANPRSWESGCTIMAPGTANSVEVFETS	138
QY	350	VDTSDNNTNHQKI-ASATLYLTDNNGVNSLAQRNDYIVFEGDGYYYQTVDQNKANKGA	408
DB	139	V-----NDTPSESYPACKGIIYADDKSVWFDFDKVTDLFQADGLYYADY----AGNEW	190
QY	409	DNIAMRDAAHY-IEDGNGDRVLVEASTGLNYGEOBIYNWLNYGGDDAFNIKSLFRILS	467
DB	191	D---FRDPHFVITPKIGTKYTALFEQGNVAMER-----GTVAVGEEIGFPVP	233
QY	468	NNDKISGRATWANAAIGLK-LNKDEKNPKVAELY SPLISAPVSDIERPNVVKLNKKYY	526
DB	234	KTEPTDGAARYCAAAIGIAQALNEARTEWK--LLPLPVTAFGVNQOTERPHVVFQNGLTY	290
QY	527	LFAATRNLGRSNDNDAWNANYAVG-----DNVAMGVYADS-LTGSIYKLNDSGVWLTVASV	581
DB	291	LFTISH-----HSTYADGLSGPDGV--YGFVSENGIFGPYEPLNGSGSLV--GN	335
QY	582	PANWRRTATSYXAVPVGAKDDQVLTYSMTN-----RNGVAGKGMDSWTWASFLLQ	632
DB	336	PSSQPYQAAYSHYVM-----TNGLTSITFDITPSSDPNVYRYG-----GTFLAPTIKLE	382
QY	633	INPDNTTTVLAK-----MTNQGDWIINDSSSEN	659

Query Match	7.2%	Score 298.5;	DB 2;	Length 413;
Best Local Similarity	26.8%	Pred. No. 2.3e+08;		
Matches 108;	Conservative 79;	Mismatches 141;	Indels 75;	Gaps 23;
QY	257	EPFLDWDSSWPQDYRTQOVA--WNGGQVLVIAMGIPNQDNH----	IYLLYNKYGDNEL	310
dbb	38	DDLWLWDTPURDINGNPVSKGNWVFLSVADRNIP-WNDRHSHARIGYFYSKDGKS--	94	
QY	311	SHWKNVGPIFYNSTAYVSQWSSGAVL--NSDINSOLFTRVDTSDNNTNHQKIASAT--	366	
dbb	95	--WYVGHLLOELANSTRATBWSGGTIMAGPSRNOQVETFTST-LFDKNGVREAAVATKG	151	
QY	367	-LYLTDDNNGVSLAQVRNDYIVFEGGYYQYQTDWKATKNKGADNIAMRDAHV-IEDJNG	424	
dbb	152	RIY-ADESG-VWFKGPDQTDLFQADGLFQYNAENLWN-----FRDPHFVFNPEGD	202	
QY	425	DRYLVFRASTGLENYQGEDQIYNWNLGYGGDAAFNKSLFRILSDNDIKSRATWANAAGI	484	
dbb	203	ETYALFEAN--VATVRGEDDI-----CED-----EIGPVPNTVVPKDNALCSASIGI	248	
QY	485	LKLNKDKNPKVAELSPISAPHYSDETERPNVVKLGNKYVLPFAATRLNRGSNDDAWN	544	
dbb	249	ARCLSPDRTE--WELLPLLTAFGVNDQMERPHVFQNGLYLETISH-----D	295	
QY	545	ANTAUG--DNVAMGYGAPDS-LTGSYKPLGDSGVYLTASVPANRRTATYSYAYVPAGKD	601	
dbb	296	STYADGLTSGDGLYGFVSGEFGFYPEPLNGSLVLGG--PASQPTAYAHYIM-----	347	
QY	602	DOYLVTSYMTN---RNG--VACKGMDSTWAPSFLLQINPDNT	638	
dbb	348	NGLVESFNEIIDRSGKVIAGSL-----APTVRVELQGHET	386	

RESULT 8  
JC2519  
levansucrase (EC 2.4.1.10) precursor - zymomonas mobilis  
N:Alternate names: sucrose 6-fructosyltransferase  
C:Species: zymomonas mobilis  
C:Date: 13-Jun-1995 #sequence\_revision 14-Jul-1995 #text

C:Accession: JC2519; PC2376  
 R:Kyono, K.; Yanase, H.; Tonomura, K.; Kawasaki, H.; Sakai, T.  
 Biosci. Biotechnol. Biochem. 59, 289-293, 1995  
 A:Title: Cloning and characterization of Zymomonas mobilis genes encoding extracellular  
 A:Reference number: JC2519; MUID:95218269  
 A:Accession: JC2519  
 A:Molecule type: DNA  
 A:Residues: 1-423 <KYO>  
 A:Cross-references: DDBJ:D17524; NID:9809529; PIDN:BAA04475.1; PID:9809530  
 A:Molecule type: Protein  
 A:Residues: 1-19 <KY2>  
 C:Genetics:  
 C:Gene: sucZE2  
 C:Keywords: glycosyltransferase; hexosyltransferase  
 F:1-19/Domain: signal sequence #status predicted <SIG>  
 F:20-423/Product: levansucrase #status predicted <MAT>

Query Match 7.2% Score 297; DB 2; Length 423;  
 Best Local Similarity 26.4%; Pred. No. 2.9e-08;  
 Matches 119; Conservative 75; Mismatches 135; Indels 102; Gaps 23;

QY 250 DAQNTTIEPLD-----VDSWVPQVVRTGOVANNNGYQVLIAMM-----GIPN 292  
 Db 24 DDPATMTTIDYDPIMTDKYWWDTPLRDI--NGQVVSFGQSVIFALVADRKYGWHN 82  
 QY 293 QNDN-HIYLLNRYGDNELSHKNGVPIFGYNSTAVSQWSSGSAVL--NSDNTSLFYTR 349  
 Db 83 RNDGARIGYFVSRGNS---WIFGHLKLDGANGPRNSWSSGCTIMAPGTANSVEVFTS 138  
 QY 350 VDTSDNTNTHQKI-ASATLTLTNGNNGVSLAQVRNDYIVFEGDGYIYQTYDQWKATKGA 408  
 Db 139 V-----NDTPSESPPAQCKGIYADKSWPFGDKVTDLFQADGLIYADI-----AENFW 190  
 QY 409 DNTAMRDHY-IEDGNGDRYLVEASTGLNNGYQEDQIYNWLYGGDDAFNKLFRILS 467  
 Db 191 D---FRDPHVFINDPDKTYALEGNNVQVR-----GAVAVGEEIEIGPVPP 233  
 QY 468 NDDIKSRATWANAAGILK-LNKDEKNPKVAELYSPLISAPMVSDEIRPNVVKLGKNGY 526  
 Db 234 KTTTPGCARYAAAGIAQALNEARTEWK---LLPLVATFAGVNDQTERPHVVFQNGLY 290  
 QY 527 LFAATRLNRGSDNDWANNANYAVG---DNVAMGVVADS-LTGSYKPLNDGSGWLTASY 581  
 Db 291 LFTISH-----HSTYADGLSGPDGV---YGFVSENGIFGPEPLNGSLVL--GN 335  
 QY 582 PANWRTATSYAVPVAGKDDQVLTYSMTN-----RNGVAGKGMDSWAPSFLLO 632  
 Db 336 PSSQPPQAYSHYVM-----TNGLVTSFDTIPSSDPNRYRG-----GTLAPTVKLE 382  
 QY 633 INPDNTTTLVLAQ---MTNOGDWINDSSSEN 659  
 Db 383 LVGHRSFVTEVKGYGYPQIEWLAEDSSN 413

RESULT 9  
 S47527  
 extracellular sucrose - Zymomonas mobilis  
 C:Species: Zymomonas mobilis  
 C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 17-Mar-1999  
 A:Accession: S47527  
 R:Song, K.B.; Lee, S.K.; Joo, H.K.; Rhee, S.K.  
 Biochim. Biophys. Acta 1219, 163-166, 1994  
 A:Title: Nucleotide and derived amino acid sequences of an extracellular sucrose gene (s  
 A:Reference number: S47527; MUID:94368848  
 A:Accession: S47527  
 A:Molecule type: preliminary  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-413 <SON>  
 A:Cross-references: EMBL:L08094

Query Match 7.1% Score 291.5; DB 2; Length 413;  
 Best Local Similarity 26.4%; Pred. No. 5.3e-08;  
 Matches 106; Conservative 77; Mismatches 147; Indels 71; Gaps 22;  
 QY 257 EPLDWDWSPVQDVRTGOVA--NWNGYOLVIAMGMPQNDNH-----IYLLNRYGDNEL 310  
 Db 38 DDLWMDTWPLRDINGNPVSFKGNWVIFSFVADRNP-WNDRHSHARIGYFSGDKS-- 94  
 QY 311 SHKNGVPIFGYNSTAVSQWSSGSAVL--NSDNTSLFYTRVDTSDNTHQKIASAT-- 366  
 Db 95 --WVYGGHLLQESANTRTAESGTTIMAPGSRNQVETFTST-LFDKNGVREAAVATKG 151  
 QY 367 -LYLTDNNGVSLAQVRNDYIVFEGDGYIYQTYDQWKATKGNADNIAMRDHY-IEDGNG 424  
 Db 152 RIY-ADSEG-VWFKGFDQSTDLFQADGLFYQYAEENLRN-----FRDPHVFINDP 202  
 QY 425 DRYLVFEASTGLNNGYQEDQIYNWLYGGDDAFNKLFRILSNDNDIKSRATWANAAGI 484  
 Db 203 ETYALFEAN--VATVRGEDI-----GED-----EIGPVANTVVPKDNLCSSASIGI 248  
 QY 485 LKLNKDEKNPKVAELYSPLISAPMVSDEIRPNVVKLGKNGYLYFAATRLNRGSDNDWANN 544  
 Db 249 ARCLSPDRT--WELGPPLLTAFGVNDQMERPHVIFQNGLYFTISH-----D 295  
 QY 545 ANYAVG--DNVAMGVVADS-LTGSYKPLNDGSGWLTASYVPANWRTATSYAVPVAGKD 601  
 Db 296 STYADGLTSGDGLYFVSENGIFGPEPLNGYGLVGG--PASQPTAEVAHYIM----- 347  
 QY 602 DQVLTYSMTN-----RNGVAGKGMDSWAPSFLLOINPDNT 638  
 Db 348 NNGLVSEFINEIIDPKSGKVRAG--GSLAPTVRVLEQGHET 386

RESULT 10  
 S39195  
 levansucrase - Erwinia amylovora  
 C:Species: Erwinia amylovora  
 C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 08-Oct-1999  
 A:Accession: S39195  
 R:Geier, G.; Geider, K.K.  
 submitted to the EMBL Data Library, September 1993  
 A:Description: Characterization and influence on virulence of the levansucrase gene  
 A:Reference number: S39195  
 A:Accession: S39195  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-415 <GEI>  
 A:Cross-references: EMBL:X75079; NID:9433558; PIDN:CAA52972.1; PID:9433559

Query Match 6.8% Score 281; DB 2; Length 415;  
 Best Local Similarity 25.6%; Pred. No. 1.9e-07;  
 Matches 109; Conservative 73; Mismatches 137; Indels 106; Gaps 22;

QY 257 EPLDWDWSPVQDVRTGOVANNNGYQVLIAMGIPN-----QNDNHIYLL-----YNYKY 306  
 Db 40 EEVFWMDTPLRDF-DGEIISVNGWCIIIFTLTADRNTDNFQODENGNYDITRDWEDRHG 98  
 QY 307 DNELSHV-KNVGP--IFGYNSTA-----VSOEWSGSVAV-LNSDNTSLFYTRVDTSDNNT 357  
 Db 99 RARICYWYSRTGKDWIFGGRVMAEGVAPTRREWAGTPIILLNDRGDIDLITYTCV-----T 152  
 QY 358 NHQKIASATLYLTDNNGVSLAQVRNDYIVFEGDGYIYQTYDQ---WKATNKGADNIAMR 414  
 Db 153 PGATIAKVRKIYTSQDSVLEGGFQVTSLSADGTTIYOTEQNAFWN-----FR 202  
 QY 415 D-AHVIEDGNGDRYLVEASTGLNNGYQEDQIY-----NWLYNGGDDAFNKLSPFR 464  
 Db 203 DPSPFIDRNDGKLYMLFEGN--VAGPRGSHETQAEWNVPPGVEDVGG----- 249  
 QY 465 ILSNDDIKSRATWANAAGILKLNKDEKNPKVAELYSPLISAPMVSDEIRPNVVKLGKNG 524  
 Db 250 -----AKYQAGCVG-LAVAKDLSSG-WQLPLPLITAVGVNDQTERPHVFPQDK 297

QY 525 YLFAATRLNRGSDNDAMNANYAVGDNV-----AMVGVYVADSLGYSKPLNDGSGVVLTA 580  
 Db 298 YLFT-----ISHKTFADNLTPGDGVYGFYSDKLTGPTPMNSGLVL--G 342  
 QY 581 VPANRTATYTYAVPVACKDQVLTSTYMTNRNGVACKGMD-----STWAPSFLLQINPD 636  
 Db 343 NPSSQPFQYSHYMP-----NGLVTSFI--DSVPKMGKDYRIGGTGTEAPVKILLKGD 393  
 QY 637 NTTV 641  
 Db 394 RSPV 398

RESULT 11  
 H90053  
 hypothetical protein fnb [Imported] - Staphylococcus aureus (strain N315)  
 C:Species: Staphylococcus aureus  
 C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
 C:Accession: H90053  
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogura, A.; Mizutani-U, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.; Lancet 357, 1225-1240, 2001  
 A:Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.  
 A:Reference number: A89758; MUID:21311952; PMID:11418146  
 A:Accession: H90053  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1038 <KUR>  
 A:Cross-references: GB:BA000018; PID:g13702453; PIDN:BA043594.1; GSPDB:GN00149  
 A:Experimental source: strain N315  
 C:Genetics:  
 A:Gene: fnb

Query Match 6.3%; Score 261; DB 2; Length 1038;  
 Best Local Similarity 18.2%; Pred. No. 7.3e-06;  
 Matches 199; Conservative 151; Mismatches 301; Indels 442; Gaps 52;

QY 55 VTGSGQVSAASD---TTTTSANANASSAANTONSQV-----ASSAATSTSS 104  
 Db 27 VGMQDKEAAAEQKTTVEENGNSATDKTSETQTATNVNHIETQSYNATVTEQPSN 86  
 QY 105 AASLNTDS-KAAQ-----ENTNAKND-----TOKAAPA 134  
 Db 87 ATQVTEAPRAVAPQAPQAPANVETKKEKPOVKETTOPDNSGNORQVDLTPKVKQTQ 146  
 QY 135 NESSE-----AKNEPANNV-DSSAAKNDQSSKKYT--TAKLNKDAENYVK 180  
 Db 147 NOGTQVEVAQPRTASESKPRVTSADVAEAEASDVSEVKGDTYTSKYTVESGSEAP 206  
 QY 181 AG--IDPNSLTDQIKALNMFNSKAASKSTQMTYNDFOKIAITLIKQGRYVVPFF-KA 237  
 Db 207 QGNKVEPIA--QQRVLYKLFADGLKRG---DYDFD-----TLSNNYNTYGVSTARKV 256  
 QY 238 SEIKN-----MPAATTKDAQNTTIE-PLDVMSNPVQ---DVRTGOVANNGYOL 283  
 Db 257 PEIRKGSVMATGILGNINRTFTNELEHKVEVTANLEINLFDPKTVQS---NGEQK 313  
 QY 284 VIAMH-----GINQ--NDNHLYLLNKKYGDNELSHKRWKVGPIFGYNSTA 326  
 Db 314 ITSKLNEETETKPIPVYNGVSNSTYNGSIETFNK-ESNKFTHIAYIKPMNGNQNT 372  
 QY 327 VSQWMS---GSVNLSDNSIQLF-----YTRVDTSDNN--TNHOKIASATLYLT 370  
 Db 373 VSVTGTITGSLNLAGQPTVKYVEILGKDELQPSQVYANTSDTNKFKDVTREMGKLSVQ 432  
 QY 371 DNNQ-NVSLAQRNDYIVVEGCGYTYQYVDQKATNKGADNIAIRDAHVIEDGNGDRVLV 429  
 Db 433 DNGSYSLNDKLDKTKVI-----HYTGEYLQ-----GSDQVNR----- 466  
 QY 430 FEASTGLENYGEDQIYNWNLNGGDAFNKSLPRILSNDIKSRATWANAAGILKLNK 489

Db 467 ---TELYGYP-ERAYKSYVYVG-----YRLTWDNGL--VLYSNK 500  
 QY 490 DEKNPKVAEL-----YSLISAPMVSDIEIRPNVYKLGNYLYFAATRLNRGS----- 537  
 Db 501 ADONGKNGQIIONDFEYKEDTAKGTMSQYDAKQIE-----TEENQDWTPLDI 550  
 QY 538 -----NDDAMNANY--AVGNVAMVG----- 557  
 Db 551 DYHTAIDGEGVYDGYVETIETDSADIDHTAVDSEVGHVGYTSEESNPIDPEE 610  
 QY 558 -----YVADSLTGSYKPLNDGTV-----LTASVPANNRATISY 593  
 Db 611 STHENSKHHADVVEEDTNPGGQVTTESNLVEFDEESTKGI VTA VSDHTTIEDTKEY 670  
 QY 594 AV-----PVAGKDDOVLVTSYMTNRNGVACKGMDSTWAPSFLLQINPDN 637  
 Db 671 TTESNLIELVDELPEHGOAGPIEETENNHHISHGLGTENGHNY--GVIEIEENS 728  
 QY 638 TTVVLAKM-----TNOGDWINDSDENLDM-----IGDLDSAAALP----- 672  
 Db 729 HVDIKSELGYEGGNSGNSQSEEDTEE-DKPKYEQGGNIVDIDFDSVPQIHGKNGDQSF 787  
 QY 673 ---GERDKP-----VDMDLIG--YGLKPHD----- 692  
 Db 788 EEDTEKDKPKYEHGNGNIIDFDSVPQIHGFKHNEIIEEDTKDKPNYQFGHNSVDFE 847  
 QY 693 -----PATNDPET--PTTPTETPTNTPKTPKTPEN 722  
 Db 848 EDTLPKVGSGNEGQOTTEEDTPTPTPTPEVPSEPTMPPTPEVPSEPTP-TPPTPEV 906  
 QY 723 PGTPPTPTNTPTPEPL-----TPETPKQETOT----- 751  
 Db 907 PSEPTP-TPPTPEVPSEPTPTPTPEVPSEPTPTPTPEVPSEPTPTPEVPSEPTPTPEVP 965  
 QY 752 -----NNRLPQTG--NNANKAMIGLGMGTLLS 776  
 Db 966 PSKPVQGVVTPVIEINEKVKAVAPTAKAOKSKSELPETGGEESTNKGML---FGGLFS 1022  
 QY 777 MGLAEINRRFN 789  
 Db 1023 ILGLALLRRNKN 1035

RESULT 12  
 A32192  
 fibronectin-binding protein - Staphylococcus aureus  
 C:Species: Staphylococcus aureus  
 C>Date: 31-Jul-1989 #sequence\_revision 31-Jul-1989 #text\_change 16-Feb-1997  
 C:Accession: A32192  
 R:Signaes, C.; Raucel, G.; Joensson, K.; Lindgren, P.E.; Anantharamaiah, G.M.; Hoeoe Proc. Natl. Acad. Sci. U.S.A. 86, 699-703, 1989  
 A:Title: Nucleotide sequence of the gene for a fibronectin-binding protein from Stap  
 A:Reference number: A32192; MUID:89098998  
 A:Accession: A32192  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1018 <SIG>  
 A:Cross-references: GB:J04151  
 C:Keywords: fibronectin binding

Query Match 6.2%; Score 255; DB 2; Length 1018;  
 Best Local Similarity 18.4%; Pred. No. 1.5e-05;  
 Matches 195; Conservative 146; Mismatches 322; Indels 398; Gaps 50;

QY 55 VTGSGQVSAASD---TTTTSANANASSAANTONSQV-----ASSAATSTSS 104  
 Db 27 VGMQDKEAAAEQKTTVEENGNSATDKTSETQTATNVNHIETQSYNATVTEQPSN 86  
 QY 105 AASLNTDSKAAQENTTAK--NDDTKAAPANESSEAKNEPANNVNDSSAKNDQDSS 162  
 Db 87 ATQVTEAPRAVAPQAPQAPANVETKKEKPOVKETTOPDNSGNORQVDLTP 146

```

QY 163 KKNTAKLNKDAENVVKKAGIDNSLTDDOIKALNKWNFSKA-----AKSGTQMT-----212
Db 147 KKAQ-----NQVAETQVEVA---OPRTASESKPRVTRSADVAEAKESAKNAVETGTDVTSKV 201
QY 213 -----YNDFOKTIAD-----TLIK-----ODGRYTVPPF-----K 236
Db 202 TVEIGSIEGHNTNKYPHAGORAVLYKYLKFKENGLHOGDYDFTLNSNVNTHGVSTARK 261
QY 237 ASEIKN-----MPAATTKDAQTNIE-PLDWTSWPVO-----DVRTGOVANNQYQ 282
Db 262 VPEIKNGSVVMATGEVLGGKTRYFTFDIEDKVIDVTAELINLFIDPKTVOT---NGNQ 318
QY 283 LVTAMM-----GIPN--QNDNHIYLLNKYGDNELSHWKNVGP1FGVNS- 324
Db 319 TITSTLNEBOTSKELDVKYKOGIGHYANLNGSIETFNK--ANNFESHVAFIKPNNGKTS 377
QY 325 -TAVSOEWSGSAVLNSDNSIOLFTRVDTSO-----NWNHOKIASATLYLT-----D 371
Db 378 VVTGTGLMKSGNONGNPKVRIEYVLGNNEIDIAKSVANTTDTSKFREVTSNMSGNLNQ 437
QY 372 NNGRVSL--AOVRNDYIVFEGDGYIYQTYDQWKATNKGADNIAHRDAHIEDNGDRIYV 429
Db 438 NNGSVLSIENLDKTYV--HYDGEYL-----NGTDEVDERTQV---GH-----477
QY 430 FEASTGLENYOGEDQIYNWLNYGDDAFNIKLSFRILSNDDIKSRATWANAAT-----482
Db 478 -----PEQLX---KYYYDRGYTL-----TWNGLVLYSNKAN 506
QY 483 -----GILKNKDE-KNPKVAELYPLISAPMVSEIERPNVVKLCNKVYLFAATRLNR 535
Db 507 GNEKNGPIONNKFEKEDT IKETLQCYDKNLVTVVEEYDSSILDIYH---TAIDG 562
QY 536 G-----SNDAMWNANY--AVGDNYAVG-----557
Db 563 GGGYVDGYIETIETDSSALDIDYHTAVDSEAGHGYTESSESNPIDFEESTHENSKEH 622
QY 558 -----YVADSLTGVKPLNDSGV-----LTASVPANWRTATVSYAV-----595
Db 623 HADVVEEDTPGGQGVVTESNLNFEEDFEESTKGIVTGAUSDHITVEDTKKYTTESNLIE 682
QY 596 -----PVAGKDDQVLVTSYMTNRNGVAGKGDSTWAPSLQINPNTTTVLAKM 645
Db 683 LVDELPEEHGAQGPVEETKNNHHISIGLCTENGCHNY--DWIEEENSHVDIKSEL 740
QY 646 -----TNOGDWTDDSSNLDM-----TGDLDLSAALP-----GERDK 677
Db 741 GYEGQNGSGNOSFEDTEE-DKPKYEGGNIVDIDFDSVPOIGHQNGKNGOSFEDTEKDK 799
QY 678 P-----VDMDLIG--YGLKPH-----691
Db 800 KYEHGGNIIDFDSVPHIGHFNKHTIEEDTNKDOKPSYQGGHNSVDFEEDTLPKVS 859
QY 692 -----DPATPNDEPPT--TPETPETPTPKTPKTPNGP 726
Db 860 QONEGQOITIEDTTPPIVPPTPTPEVPSEPETPTPTPEVPSEPETP--TPPTPEVPSE 918
QY 727 QTPVNTNPTELTP-----ETPKQPE-----750
Db 919 ETP--TPPTPEVPAEPGKVPVPAKEEPKSKPVQEGKVTVPTVIEINEKVAVPTKKPOS 977
QY 751 TNNRLPOTG--NNANKAMICLNGTLLSMFGLAEINKRRFN 789
Db 978 KXSELPETGGEESTNKGML---FGGLFSILGLALLRRNKN 1015

```

RESULT 13

S19702  
fibronectin-binding protein B - *Staphylococcus aureus*  
C:Species: *Staphylococcus aureus*  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text  
C:Accession: S19702  
R:Joensuu, K.; Signaes, C.; Mueller, H.P.; Lindberg, M.

Eur. J. Biochem. 202, 1041-1048, 1991  
A>Title: Two different genes encode fibronectin binding proteins in *Staphylococcus aureus*  
A/Reference number: S19702; MUID:92111475  
A/Accession: S19702  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-940 <JOE>  
A/Cross-references: EMBL:X62992; NID:949040; PIDDN:CAA44726.1; PTD:9581562  
C/Keywords: fibronectin binding

Query Match 6.0%; Score 246; DB 2; Length 940;  
Best Local Similarity 18.6%; Pred. No. 3.8e-05;  
Matches 185; Conservative 135; Mismatches 334; Indels 340; Gaps 44;

	Conservative	135:	Mismatches	334;	Indels	340;	Gaps
QY	55	VTLGSG-----VSAASDTTIRTSANANSASSAANTONSNOVASSAAITSSSTSSAASLN	109				
Db	25	IIVGMGOEKAASEONNTTVESGGSATSKASETQTITTNNVT---IDETOSYSTAT81		:	:	:	:
QY	110	NTDSKAAQENINTAKNDDTKAAPANESESEAKNEPAVNND---SSAAKNDDQOSS----	162				
Db	82	EOPSQSQTAVTEEPK--TVQAPKEISRV-DLPSEKVADKETTTCTOVDIAQPSNVSEI	137	:	:	:	:
QY	163	----KNT--TAKLNDAENVVKAGIDPNLSLTD-----DQLK	194	:	:	:	:
Db	138	KPRKRSTDVTAVEREEVEETKATGTDVTNKVEVEEGSEIVGHKODTNVNPNAERT	197	:	:	:	:
QY	195	ALNMNFSSKAASKGTOMTYNDFOKIADTLIKODGRVTPFPFKASEIKNNPAAATDKAOIN	254	:	:	:	:
Db	198	LKYKWFGEIGKAG--DYDFE-TLSD-----NVETHGISLRKVPKISTGDGV-	243	:	:	:	:
QY	255	TIEPLDVMSWPQDVRTGOVAWN-----GYOLVIANMGIPNON	294	:	:	:	:
Db	244	----MATGEIIGERKVRVTFKEYVOEKDLTAELSLNFLDPITVTQK	288	:	:	:	:
QY	295	DNHIIYLLNYKCYDNELSHWKNVPIFGYNSTAVSOEWSSGSAVLNSDNSIQLFYTRVD-	352	:	:	:	:
Db	289	NONVEZ-----KLGETTVSRIFNIQVLGG-----VRDNMGVTA-----NGRIDTLN	329	:	:	:	:
QY	353	-SDNNTNHOKIASATLYLTDNNGNVSLAQVRNDYIVPEGDYYVOTYDQWKATNKAGANI	411	:	:	:	:
Db	330	KVDGFESH-----FAYMKPNQSSLSSVTVTGQTKGNKGVNNPTVKVTK--HIGSDOL	381	:	:	:	:
QY	412	A-----MRDAHVIEDGNGRYLVFEASTGLE-NYOGEDOIYNWL-----NYGDDAFNIK	460	:	:	:	:
Db	382	AESVYAKLDVSKFEDVTDNMSLDFDTNGYSFLNFNLDSKNVVIKEYGYDSNASLE	441	:	:	:	:
QY	461	SLEFILS--NDDTKSRATAWAAAIGILKNKDKNPKVAELYSPLI--SAPVUSDIERP	516	:	:	:	:
Db	442	FQTHLFGYINY YTSNLTKWN--GVAFYSNNAQGCKOKLEPIIHEHTPIELEPFSEP	498	:	:	:	:
QY	517	NVYKLGKKYLYFAATRLNRGSSNDAMNANY-----AVG-----550		:	:	:	:
Db	499	PVEK-----HELTCG I EESNDKSIDFEYHTVABEGAEGAGTETEEDSIHVDFEES	551	:	:	:	:
QY	551	-----DNVAMYCVADSLTGSYKPLUNDGUV-----LTASVPANNWTATSYVA	594	:	:	:	:
Db	552	THENSKHADVYEIEDTNPGGOGVOTESNLVEFDDEDSTKGIVTCAVSDHTTIETDKYT	611	:	:	:	:
QY	595	V-----PVAGKDDQVLVITSYMTNRNGVAGKMDSTWAPSFLLQINPONT	638	:	:	:	:
Db	612	TESNIELVELPEEHGAOQP IEBITENNHHISHGLGTENGHCNY--GVIEIEENS	669	:	:	:	:
QY	639	TTVLARK-----TNOGDWIWDSSSENLM-----IGDLDSAALP-----	672	:	:	:	:
Db	670	VDIKSELVEYGONGSONSFEDTEE-DKP KYEQGNIVDIDFDSVPOIHGONGNQSF	728	:	:	:	:
QY	673	--GERDKP-----VDWDLIG--YGLKPH-----	691	:	:	:	:
Db	729	EDTEKDKPKYEQGNIIDIDFDSVPHIGHFNKHFTIEEDTNKDKPNYQFGHNSVD FEE	788	:	:	:	:
QY	692	-----DPATPDNETPTTPEPTPNTKPTKTENPGTPTPTN	733	:	:	:	:

Db	789	DPLPOVSGHNGQQTIEDTPPIVPPTPTPEVSEPEP-TPTPEVSEPEP-TTP	846
QY	734	TPETPLTP-----ETKQPE-----	
		:	
Db	847	TPETPEPGPKPIPAKEEKPKSPKEQGVVTPVIEINEKAVVPTTKAQSKEP	757
QY	758	TC--NNANKAMIGLGMGTLLSMFGLAEINKRRFN	906
		:	
Db	907	TGGEESTNNGML---EGGLFSILGALLRRNNKN	937

RESULT 14  
 FC50AG  
 IGA FC receptor precursor - Streptococcus agalactiae  
 N/Alternate names: beta antigen  
 C/Species: Streptococcus agalactiae  
 C/Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 16-Jul-1999  
 C/Accession: S15330, S20240; S17038  
 R/Jerlstroem, P.G.; Chhatwal, G.S.; Timmis, K.N.  
 Mol. Microbiol. 5, 843-849, 1991  
 A/Title: The Iga-binding beta antigen of the c protein complex of Group B streptococci:  
 A/Reference number: S15330; MUID:91312121  
 A/Accession: S15330  
 A/Molecule type: DNA  
 A/Residues: 1-1164 <JERI>  
 A/Cross-references: EMBL:X59771  
 A/Accession: S20240  
 A/Molecule type: protein  
 A/Residues: 38-48 <JEF>  
 R/Jerlstroem, P.G.  
 submitted to the EMBL Data Library, August 1991  
 A/Reference number: S17038  
 A/Accession: S17038  
 A/Molecule type: DNA  
 A/Residues: 1-914, 'E', 916-1164 <JEB>  
 A/Cross-references: EMBL:X59771; NID:g46522; PIDN:CAA42442.1; PID:g46523  
 C/Superfamily: IGA FC receptor  
 C/Keywords: cell wall; immunoglobulin receptor; tandem repeat; transmembrane protein  
 F/1-37/Domain: signal sequence #status predicted <SIG>  
 F/38-1164/Product: IGA FC receptor #status experimental <MAT>  
 F/199-438/Domain: IGA binding #status predicted <IGAI>  
 F/439-827/Domain: IGA binding #status predicted <IGA2>  
 F/827-945/Region: proline-rich repeats  
 F/946-1131/Domain: cell wall-spanning #status predicted <CWS>  
 F/1132-1159/Domain: transmembrane #status predicted <TMW>

[illegible]

```

Db      314 K-NOFLNAQKLEIOPLIKET---NVKLYKAMSESLEQVEKELKHSEANLEDLVAKS 369
Qy      257 EPLDWDVS-----WPDQVTRTGOVAN 277
Db      370 EIVREYEGKLNQKNLDELKOLEEAEHSHKLQVVEDFRKKFKTSEQVTTPKKRYKRDLAAN 429
Qy      278 WNGYQ----LVITAMGIPINQNDNHILLYLNKYGNDELSHKWNPGIFGYNTAVSQWSGS 334
Db      430 ENNOOKVELTVSPENITVIEGEDVKFTVTAKSCKTT--LDPSDLLTKNYPNSDRISTN 487
Qy      335 AVLNSDNSIQFLTFRVDTSNNHTNHOKIASATLYLTLDNNGNVSLAQVRNDYIYFEGGGYY 394
Db      488 YKTNTDN---HKIAEITIKNLKNESTVTTLKAKDDSGNV---VEKTFITV-----533
Qy      395 QTYDQWKATNKAGDIAMRDAHVIE-----DNGDRILVFPASTGLENYOGEDOI 445
Db      534 -----QKKEBKQVPKTEQKSKTEBKPQEPKSDNDKQLQELIKSAQOEKLE-----583
Qy      446 YNWLNYYGGDDAFNIKSLPRILSNDIOKS-----RATWAN-----AAIGILINKDE 491
Db      584 -----RAIKELMQPEIPSNEYGIQKSIWESQEPIQEAITSFKKIIGDS 629
Qy      492 KNPKVABLYSPLISAPWYSDETERP-----NVVKLGHNXYLFPAATRNRGNSDDAWMNA 545
Db      630 SSKYYTEHYFNKYKSDENMYQLHAQWEMLTRKVQVMHMKYPDNAEIKKFIFESDMRKRED 689
Qy      546 NYAVGDNMVAMYGVADSLTGSKYKL-----NDSGVVLITASVPANWRRTATSYIYAVPVACKD 601
Db      690 NYGSLDENALGJEFKYFLTPFNKIKQIVDDLDKVEQDQA-----PIPENSEM 739
Qy      602 DQ-----VLVTSYMTNRNGVAGKMGDMSTWAPSFLLOINPDNTTVVIKAKMTNQGDWIWDD 655
Db      740 DQAKEKAKIAVSKZMS-----KVLQGVHQ-----HLQKKNNISKIV-----DLPEKL 780
Qy      656 SSLENLMIGLDSLAAALGERDKPVDMDLIG-----YGLK---PHDPATPNDETPT 702
Db      781 EAIKQOTIFDIDNAKTEVEDNLVH-DAFSKMNATVAKFOKLETNTPEYDTPKXIELP 839
Qy      703 TTPPETPTNPKTPTKENPGTPTENTNTPEILPTBETPKQEPQTNNRLPQT 758
Db      840 QAPDTQPAPTTHPVESPKAPEAPRVPESEKTPTEAPHVPESEKAPRVPESEKPT 895

```

RESULT 15  
A60234  
Iga Fc receptor precursor - Streptococcus agalactiae (strain SB35)  
N:Alternate names: Iga-binding protein; protein Bac  
N:Contains: beta antigen  
C:Species: Streptococcus agalactiae  
C:Date: 08-Dec-1992 #sequence\_revision 08-Dec-1992 #text\_change 26-Aug-1999  
C:Accession: A60234; S14595; A60230  
R:Heden, L.O.; Frithz, E.; Lindahl, G.  
Eur. J. Immunol. 21, 1481-1490, 1991  
A>Title: Molecular characterization of an Iga receptor from group B streptococci: se-  
ents with Iga-binding capacity.  
A:Reference number: A60234; MUID:91257158  
A:Accession: A60234  
A:Molecule type: DNA  
A:Residues: 1-1134 <HED>  
A:Cross-references: EMBL:X58470; NID:g46520; PIDN:CAA41384.1; PID:g46521  
R:Heden, L.; Frithz, E.; Lindahl, G.  
submitted to the EMBL Data Library, March 1991  
A:Description: Molecular characterization of an Iga receptor from group B streptococ-  
fragments.  
A:Reference number: S14595  
A:Accession: S14595  
A:Molecule type: DNA  
A:Residues: 1-1134 <HE2>  
A:Cross-references: EMBL:X58470; NID:g46520; PIDN:CAA41384.1; PID:g46521  
R:Lindahl, G.; Akerstrom, B.; Vaerman, J.P.; Stenberg, L.  
Eur. J. Immunol. 20, 2241-2247, 1990

A:Title: Characterization of an IgA receptor from group B streptococci: specificity for  
 A:Reference number: A60230; MUID:91055597  
 A:Accession: A60230

A:Molecule type: protein

A:Residues: 'X',39-48,'X',50-52,'X',54-56 <LIN>

C:Superfamily: IgA Fc receptor

C:Keywords: cell wall; immunoglobulin receptor; tandem repeat; transmembrane protein

F:1-37/Domain: signal sequence #status predicted <SIG>

F:38-1134/Domain: IgA Fc receptor #status experimental <MAT>

F:199-438/Domain: IgA binding #status predicted <IGAL>

F:439-826/Domain: IgA binding #status predicted <IGAL>

F:827-915/Region: proline-rich repeats

F:916-1101/Domain: cell wall-spanning #status predicted <CWS>

F:1102-1129/Domain: transmembrane #status predicted <TMV>

#### Query Match

Best Local Similarity 5.8%; Score 237.5; DB 2; Length 1134;  
 Matches 189; Conservative 129; Mismatches 349; Indels 273; Gaps 39;

Qy	6	KMAVVTLSAALVFGATTNASADTNIENDSST-----VOVTTGDNDAIAKSVTLGS	59
Db	42	KDSDVKTTEVAAPY-----PSAQTQGNSSSSSELETTKMEIPTDIKRAVEPVETA	96
Qy	60	GOVSAASDITRTSA-----NANSASSAANTONSQVASSAITSSTSSASL-----	108
Db	97	GETS-ATDTGKREKQLQWKNLKNVDNTILSHQKNEFKTKIDETNDSALLELENOF	155
Qy	109	NNT-----DSKAAOENTNTAKNDTQ-----KAAPANESSEAK- 141	
Db	156	NETNRLHLHIQHEEVEKKAQKQ--KTLKQSDTKVDLSNIDKELNHQKSOVEKMAEQK	213
Qy	142	--NEPAVNVNDSAAKND-----QSSKK-----NTAKLNKDAENVV	178
Db	214	ITNED-----KDSMLKIEDIRKQAQADKDAEVKVEELKGLFSSTKAGLQOEIHEV	269
Qy	179	KK-----AGID---PNSLTDQIKALKNFSSKAAKSQTMNTYNDF-----OKIAD--	221
Db	270	KKETSSEENTQKVDHEANSQNLAKSLKLELDKATTEQATQVYK-NOFLKNAQKLEIQ	328
Qy	222	TLIKODGRYTPFFKA-----SEIKNMPAATTKDAQTNTIEPLDYVDS-----	264
Db	329	PLINKT-----NVKLYKAMESLEQVEKELKHNSANLEDLVAKSKEIVREYEGKLNQSKNL	385
Qy	265	-----WPVQDVRTGQVAVNWNGYQ---LVIAMMGI 290	
Db	386	PELQLEEEAHSKLQVVEDEFKFKTSEQVTPKKRKRLDAAANNENQOKIELTVSPENI	445
Qy	291	PNQNDNHIIYLLYKNGDNLHWNKNGPIFGYNSTAVSQEWSGSAVLNSDNLQFLPYTRV	350
Db	446	TVYEGEDVKTVTAKSDSKTT--LDFSLLTKYNPSVSDRISTNYKNTDN-----HKIA	498
Qy	351	DTSDNTNHQIKASATLYLTNDNNGVSLAQVRNDYIVFEGDGYQTYQYDQWKATNKGADN	410
Db	499	EITIKNLKLNESQTVTLAKKDDSGNV-----VEKFTIV-----OKKEEQVPKT	544
Qy	411	IAMRDAHVLE-----DGNDRYLVFEASTGLTNGEDQIYNWLNLYGDDAFNIKS	461
Db	545	PEQDSKTEEKVPQEPKSNKDLQQLIKSAQOELEKLE-----KA 585	
Qy	462	LFRILSNDDIKS-----RATWAN-----AAGILKLNKDEKNPKVAELYSPLISAP	507
Db	586	IKELMEQPEIPSNPEYGTQKSIWESQKEPIQEAITSFKKIIGDSSSKYYTEHYENKYKSD	645
Qy	508	MYSDEIERP-----NVKVLGNKYLFATRLNRGSDNDAMNANYAVGNVAMVGVAD	561
Db	646	FNYYQLHAQMEMLTRKVVQYNNKYPDNAEIKKIFESDMKTKEDNYGSLNDALKGYFEK	705
Qy	562	SLTGSYKPL-----NDSGVVLFTASVPANWRATYSYAVPVAGKDDQ-----VLVTSYMT	611
Db	706	YFLTPFNKIKQIVDDLKKVQDQPA-----PIPENSEMDOAKEKAKIAVSKYMS	755
Qy	612	NRNGVAGKGMDSWAPSFLLIQINPDNTTTLAKMTNOGDWIDDSSENLDNIGDLSAAL	671

Search completed: September 26, 2002, 18:25:06  
 Job time: 85 sec

Db	756	-----KVLGCVHQ-----HLQKKNHKIV-----DLKLELEAIKQQTIFDIDNAKT	796
Qy	672	PGERDKPVDWDLIG-----YGLK---PHDPATNDPDTPTPTPTPTPTPTPTPTPK	718
Db	797	EVEIDNLVH-DAFSKMNATVAKFQGLENTPTETPTDKIPELQAPDTQAPDTPHVE	855
Qy	719	TPENPCTPQTPNTPTNTPELPTPTPKQPTQTNNRLPQT	758
Db	856	SPKAPAPRVPEPKTPPEAPHVPEPKTPPEAPKIPPEPKT	895



Result No.	Query #	Score	Query			ID	Description
			Match	Length	DB		
1	798	19.3	881	22	ABG21651	Novel human diagno	
2	798	19.3	893	22	ABG07167	Novel human diagno	
3	793.5	19.2	1027	22	ABG21645	Novel human diagno	
4	766	18.6	736	22	ABG32589	Novel human diagno	
5	760.5	18.4	774	22	ABG05590	Novel human diagno	
6	760.5	18.4	774	22	ABG11754	Novel human diagno	
7	759.5	18.4	789	22	ABG04946	Novel human diagno	
8	759.5	18.4	789	22	ABG28407	Novel human diagno	
9	759.5	18.4	1095	22	ABG21602	Novel human diagno	
10	759.5	18.4	1095	22	ABG21823	Novel human diagno	
11	758.5	18.4	527	22	ABG04307	Novel human diagno	



QY	561	DSLTCGSYKPLNDGCVLTASPVANWRTATVSYVAVPVAGKQDQVLVTSYMTNRNGVACKG	620
Db	791	nslgtgpgkplnktglvlqmgldpndvtfysfhavpqa-kgnvvvitsymtnrgffedk-	848
QY	621	MDSTWAPSFLIQNPQNTTTLAKMTNQG	649
Db	849	-katfapsflmnlkgnktsvvnksileqg	876
RESULT 2			
ID	ABG07167	standard; Protein; 893 AA.	
XX	ABG07167;		
DT	13-FEB-2002	(first entry)	
XX			
DE	Novel human diagnostic protein #7158.		
XX			
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;		
KX	food supplement; medical imaging; diagnostic; genetic disorder.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200175067-A2.		
XX			

30-MAR-2001; 2001WO-US08631.  
31-MAR-2000; 2000US-0540217.  
23-AUG-2000; 2000US-0649167.  
(HYSE-) HYSEQ INC.  
Drmanac RT, Liu C, Tang YT;  
WPI; 2001-639362/73.  
N-FSDB; AAS71354.  
New isolated polynucleotide and encoded polypeptides, useful in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits and to assess  
biodiversity -  
Claim 20; SEQ ID No 37526; 103pp; English.  
The invention relates to isolated polynucleotide (I) and  
polypeptide (II) sequences. (I) is useful as hybridisation probes,  
polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
and gene mapping, and in recombinant production of (II). The  
polynucleotides are also used in diagnostics as expressed sequence tags  
for identifying expressed genes. (I) is useful in gene therapy techniques  
to restore normal activity of (II) or to treat disease states involving  
(II). (II) is useful for generating antibodies against it detecting or  
quantitating a polypeptide in tissue, as molecular weight markers and as  
a food supplement. (II) and its binding partners are useful in medical  
imaging of sites expressing (II). (I) and (II) are useful for treating  
disorders involving aberrant protein expression or biological activity.  
The polypeptide and polynucleotide sequences have applications in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits to assess biodiversity  
and to produce other types of data and products dependent on DNA and  
amino acid sequences. ABG00010-ABG30377 represent novel human  
diagnostic amino acid sequences of the invention.  
Note: The sequence data for this patent did not appear in the printed  
specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences.  
Sequence 893 AA;

Query Match 19.3%; Score 798; DB 22; Length 893:

		Best Local Similarity	34.4%; Pred. No. 4.7e-41;	
Matches	237; Conservative	119; Mismatches	253; Indels	Gaps
QY	7 NNAVVTLSAA-LVPGATTWASAD-----TNIENND-SSTVVQTGDNDIAVKSVTIGS	59		
Dg	234 dydvividsapnlgitlnnvcaadylvptpaelfdytsalqfddmldl-lknvdllkg	292		
QY	60 GOVSA-ASDITRTISA-----ANSASSAANTQNNSOV---ASSAITTSSTS-AASLNN	110		
Dg	293 fepdaetsggalgltlanvvtlsanleslngdtselqpedsaseitrafdtlakaln	352		
QY	111 TDSKAAOENTWTAKNDDTKRAAPANESSEKAKEPANNV-----NDS-----SAAKNDQQGS	161		
Dg	353 tdsgss----spslagidctsggsahistrdqstplieevgepllsdthvfksiredngr	409		
QY	162 SKR--NTTAKLNKDANVVKKAGI---DPNSLTDDQIKALNMKNFSKAASKG---	TQMTY	213	
Dg	410 sqkvhtegdmnmnikvkqatvlfttallaggatqafakenngkayketygvshlr	469		
QY	214 NDFOKIADTLTIKODGRYTVPFFKASEIKMPPATTKDAQNTIEPLVDWDSPVQDVRTG	273		
Dg	470 hdmqlipkq-qgnekyqvpgfqdstikniesa-----kgldvwdsplqna-dg	516		
QY	274 QVANWGCVQLVIAMGIP-NQDNHIIYLKYGNELSHWNKVGP1-----FCYNSTA	326		
Dg	517 tvaeyngyvhvalagspkaddtsiyymfyqvgdnstdwnagrvmfkdsdkfdandpi	576		
QY	327 V---SQEWSGSVALNSDNSIQLFYTRVDTSDNTHNQIASATLYLDNNGNVSLAQVRN	383		
Dg	577 lkdqtewsgsatfsdgklrlyft--dysghykqgalttaqvvnvsksdtlikingved	634		
QY	384 DIIVFEDGGYYTYTDQW--KATNKGANIAMDAHVIDEGNGDRYLVEASTGLEN-YQ	440		
Dg	635 htifdgdgktynqvqqfidegnysgtndhtirdphyyed-kghkylveantgtengyv	693		
QY	441 GEDOYINWLNGDGDDAFNIKSFILSNDDIKSRATWANAAIGILKLNDKPNKVAELY	500		
Dg	694 geesinfkay999gtnfrkesqk-lqqsakkrdaelangalgieleindytikkvmk-	750		
QY	501 SPLISAPMVSEDIERPNVVLGNKYFLFAATRLNRGNSDDAWMNANYAVGDVNVMGYVA	560		
Dg	751 -Plitsntvtdeieranvfmngkwylftrsrgskmtidg-insn-----diymlygyvs	802		
QY	561 DSLTGVSKPLNDSCVGLTASVPANWRATYSYYPVAGDKDOVLTVSYTNRNNGVAGKG	620		
Dg	803 nsltpgpypinktgivilqmglpndvtfyshfavpqa-kgnvvitsymtngffedk-	860		
QY	621 MSTWAPSFLQINPDNTTTVLAKMTNOQ	649		
Dg	861 -katfapsflmnknkgtsvvksnllegq	888		
RESULT	3			
ID	ABG21645			
XX	ABG21645 standard; Protein; 1027 AA.			
AC	ABG21645;			
XX	ABG21645;			
XT	18-FEB-2002 (first entry)			
DE	Novel human diagnostic protein #21636.			
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;			
XX	food supplement; medical imaging; diagnostic; genetic disorder.			
OS	Homo sapiens.			
PX	WO200175067-A2.			
PD	11-OCT-2001.			
PF	30-MAR-2001; 2001WO-US08631.			
XX				

PR	31-MAR-2000; 2000US-0540217.			
PR	23-AUG-2000; 2000US-0649167.			
PA	(HYSE-) HYSEQ INC.			
XX	Drmanac RT, Liu C, Tang YT;			
PI	WP1, 2001-639362/73.			
DR	N-PSDB; AASU5832.			
XX	New isolated polynucleotide and encoded polypeptides, useful in			
PT	diagnostics, forensics, gene mapping, identification of mutations			
PT	responsible for genetic disorders or other traits and to assess			
PT	biodiversity			
XX	Claim 20; SEQ ID NO 52004; 103pp; English.			
PS	The invention relates to isolated polynucleotide (I) and			
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,			
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome			
CC	and gene mapping, and in recombinant production of (II). The			
CC	polynucleotides are also used in diagnostics as expressed sequence tags			
CC	for identifying expressed genes. (I) is useful in gene therapy techniques			
CC	to restore normal activity of (II) or to treat disease states involving or			
CC	a food supplement. (II) and its binding partners are useful in medical			
CC	quantitating a polypeptide in tissue, as molecular weight markers and as			
CC	imaging of sites expressing (III). (I) and (II) are useful for treating			
CC	disorders involving aberrant protein expression or biological activity.			
CC	The polypeptide and polynucleotide sequences have applications in			
CC	diagnostics, forensics, gene mapping, identification of mutations			
CC	responsible for genetic disorders or other traits to assess biodiversity			
CC	and to produce other types of data and products dependent on DNA and			
CC	amino acid sequences. ABG00010-ABG30377 represent novel human			
CC	diagnostic amino acid sequences of the invention.			
CC	Note: The sequence data for this patent did not appear in the printed			
CC	specification, but was obtained in electronic format directly from WIPO.			
CC	at ftp.wipo.int/pub/published_pot_sequences.			
XX	Sequence 1027 AA;			
SQ				
	Query Match	19.2%; Score 793.5; DB 22; Length 1027;		
	Best Local Similarity	34.1%; Pred. No. 1.e-40;		
	Matches	217; Conservative 110; Mismatches 232; Indels 77; Gaps 24;		
QY	64 AASDITRTISA-----NANSASAANTQNNSOVA--SSAITSSTSSAASLNTDSKAA	116		
Dg	182 agteactvgstvshhppnpvpvsgtkvndvsnmmysrnagtvrqmlskrvsesappsf	241		
QY	117 QENTTAKNDDTKAAPANESSEKAKE-----PAYN-----VNDSSAANKDD	158		
Dg	242 qtpvntsvsstnl-vtppvavvsqpkicptvtseltatsvipatvatvatsiqredr	300		
QY	159 QQSSKK--NTTAKLNKDANVVKKAGI---DPNSLTDDQIKALNMKNFSKAASKG---	TQ	210	
Dg	301 ngrrsqktvhtegdmnmnikvkqatvlfttallaggatqafakemqketygvsh	360		
QY	211 MTYNDFOKIADTLTIKODGRYTVPFFKASEIKMPPATTKDAQNTIEPLVDWDSPVQDV	270		
Dg	361 itrhdmlqipk-qgnekyqvpgfqdstikniesa-----kgldvwdsplqna	408		
QY	271 RTGOVANGCQLVIAMGIP-NQDNHIIYLKY			

Qy	439	-YGEDOIYNWLYNGGDAFNIKSLFRILSNDDIKSRATWANAAIGILKLNKDEKNPKVA	497
Db	585	gyggeesafnakyvyyggntffrrkesqk-lqggaakrdaelangaigliehndytlkkvkm	643
Qy	498	ELXSPLISAPWSDTERPNVYKLGHKYFLFAATRLNRGNSDDANWANNAYGDNVAMVG	557
Db	644	k---plltstvtdeleeravfmngkwyfltdrsgskmtldg---lnsn-----diymlg	693
Qy	558	YVADSTGSKYKPLNDGSGVVLTSVFPANMWTATSYAYVAVAGKDDQVLVTSYMTNRNGVA	617
Db	694	yvsnsitgpykplnktgvlqmgldpndvtfyfshfaypqa-kgnnvittsyntcrgffe	752
Qy	618	GKGMDSWTAPSFLLQINPDNTTIVLAKWTNQGDWIT	653
Db	753	dk--katfapsflmnlkgnstsvvksnllsgqgltw	786

RESULT	4	
ABG22589		
ID	ABG22589	standard; Protein; 736 AA.
XX		
XX		
AC	ABG22589;	
XX		
DT	18-FEB-2002	(first entry)
XX		
DE	Novel human diagnostic protein #22580.	
XX		
XX		
KK	Human; chromosome mapping; gene mapping; gene therapy; forensic;	
XX	food supplement; medical imaging; diagnostic; genetic disorder.	
KK		
XX		

CC	diagnostic amino acid sequences of the invention.	
CC	Note: The sequence data for this patent did not appear in the printed	
CC	specification, but was obtained in electronic format directly from WI	
CC	at ftp.wipo.int/pub/published_pct_sequences.	
XX		
SQ	Sequence 736 AA;	
	Query Match	18.6%; Score 766; DB 22; Length 736;
	Best Local Similarity	35.2%; Pred. No. 3.5e-39;
	Matches 206; Conservative 101; Mismatches 219; Indels 60; Gaps	
QY	88 SNSQVSSAAITSTSSAAS-----LANTDSKAAQENTAKNDDTOKAAPANESSEAKNE	143
Db	182 tgnmvkpkaktvdtlsagggdvnmvfdepnkeeaqvtpsd-----knktyseyvl	235
QY	144 PAVNVNDSSAAKNDQOSSFK-NYTKLNKNDKAEVNVKKAGI---DPNSLTTDQIKALNKM	199
Db	236 pkvtgllhksirednrgsrqkvthtegdmnmkikvkqatvfttallagatqafake	295
QY	200 NFSKAAKSG---TQMVTYNDFOKLTADFLIKODGRYTYPPFKASEIKMMPAATTKDAQTWI	256
Db	296 nnqkayketygvshirhdmqlpkq--dqnekyqvpqfdqstkniesa-----	343
QY	257 EPLDVTDSWPQDVRTGOVANNNGYOLVITAMGIP-NONDNHIIYLLNKKYGDNELSHWKN	315
Db	344 kgidvdwswplqna-dgtvaeyngyhvvalagspkdaddtslymfykvkgvgnslswn	402
QY	316 VGPI-----FGVNSTAV---SQEWSGSVAVLSNDSNSIQLFYTRVTDSONTNHOKIASAT	366
Db	403 agrvfksdkdfandpilkdqgewsgeatfstdgkrlifyt--dysgkhykgqslttaq	460
QY	367 LYITDNNGNVSLAQVRNDYIVFEGDGYYYQTYDQW--KATNKGADNIAMRDAHVIEDGNG	424
Db	461 vnysksddtlkngvedhktlfdgdktyqnvqgfidegnytsqgdnhtcirophyyed-kg	519
QY	425 DRYLVFEASTGLEN-YQGEDQIYNWLNTGGDDAFNIKSLFRILSNDDTKSRATWANAAGI	483
Db	520 hkyIVfeantgtengyqgeesifnkay7gggtfnfrkesqk-lqgsakkrdaelangelg	578
QY	484 ILKLNKDEKNPKVAEILSYPLISAPWSDIEIRPNVVKLGKVKYLYFAATRLNRGSDDAWM	543
Db	579 ilelnndytlkkvmk---pilisntvtdeieranvfkmngkwyfstdrgrskmtldg--i	633
QY	544 NANYAVGDNVAMVGTVAOASLTGSGYKPLNDSGVLTASTAPNWRRTATYSYAVPVAGDDQ	603
Db	634 nsn-----diymlgvyvsnltgpykplnktglvlgmqldpndvrtfytshfavpqqa-kgmn	687
QY	604 VLVTYSMTNRRNCVACKGMDSTWASFLQIINPDNTVTTVLAKMNNQ	649
Db	688 vvitssynchrgefdekk--katfapsflmllkgnktsvvnksileqq	731

RESULT	5	
ABG05590		
ID	ABG05590	standard; Protein; 774 AA.
XX		
XX	ABG05590;	
XX		
DT	13-FEB-2002	(first entry)
XX		
DE	Novel human diagnostic protein #5581.	
DE		
XX	Human; chromosome mapping; gene mapping	
KW	food supplement; medical imaging; dia	
KW		
XX		
OS	Homo sapiens.	
XX		
PN	WO200175067-A2.	
XX		
PD	11-OCT-2001.	
XX		
XX	30-MAR-2001; 2001WO-US08631.	
PF		



New isolated polynucleotide and encoded peptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.

Claim 20; SEQ ID NO 35305; 103pp; English.

The invention relates to isolated polynucleotide (I) and polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical disorders involving aberrant protein expression or biological actions in the polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of DNA and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 789 AA;

Query Match 18.4%; Score 759.5; DB 22; Length 789;  
Best Local Similarity 37.2%; Pred. No. 9, 7e-39;  
Matches 193; Conservative 90; Mismatches 187; Indels 49; Gaps 18

DQ 150 DSSAAKNDQQSKNTAKLNKDENVVKKAGI--DPNSTDDOIKALKNMNFSAKA 206  
|| || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
296 dsiredngsrsktvhtegdmnnlkkivkgatvfttailagatqafakennqayk 355

DQ 207 SG--TQMTYNDFQKIADTLIKODGRYTVPPFKASEIKMPAATTKDAQTNIPLDVWD 263  
|| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
356 etygvshltrhdmlqipkq-qgnekypvpfdqgtikulesa-----kglawd 403

DQ 264 SWPVDVARTGVANWNGYQLVTAMGIIP-NONDNHILLYNKYGDNELSHKWVGPI--- 319  
|||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
404 swplqna-dgtvaeingyhvfalagspkddattslymfykvgdnslsdwkagrvmkd 462

DQ 320 ---FGYNSTAV---SOEWSGSVAIVNSONSIOLFYTRVDTSDNNTHHOKIASATLVLTENN 373  
Db | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
463 sdkfaandplliddqtqwsgsatfstdgkirflyt-dysghkygkslttagvnyvksd 520

DQ 374 GNVLAOVRNDYIFVEGDGYYYQTPDOW--KATNKGADNIAEMRDARHYEDGNDRYLVEF 431  
Db || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
521 dtlkngvedhkdtifdgdktygnqvqfidegnystgdnhitrdphyved-kghkiylve 579

DQ 432 ASTGLENL-YGEDEQIYNLWNYGGDDAFNIKSFLFRILSNDDDIKSRATWANAAIGTLKLND 490  
Db |||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
580 anttgengyggeeslfnkayyyggntffrkesdk-lqsaksrkdaelangelniend 638

DQ 491 EKNPKVALYSPLISAPWSDEIERPNVVKLGNKYLLFAATRNRSGNSDDAWNANYAVG 550  
Db || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
639 ytlkkymk---plltantvtdeleianfvkmkgkwylftdstrgskmtldg--lnsn---- 689

DQ 551 DWNVAYGIVYADSUTSGLPKLPDLSGVVLTSVPANWTATSYTAVPVACKDQDLVTSYM 610  
Db ::||:::|||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
690 -dlmlygyvsnsltpgypklntkvlgmqldpnadvftyshfavpqa-kgnnvvtlsym 747

DQ 511 TWRNGVAGKMOSTWAPSFLQLINPDNTTTVLAKMTNQG 649  
|| || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
748 trngfedk--katfapslfmnikanpktasvsvktseia 784

Db	356	etygvshitrhmlqpkq--qqnekygvqpfadastiknlesa-----kgldvwd	403
QY	264	SNPQDVRTGOVANNNGYQIYAMGIP-NONDNIHYLLNKKYGDNELSHWKNVGP----	319
Db	404	swplqna-dgtvaeingvhwvfaIagspkdaddtsImfyqkvgdnsidswknagrvfk	462
QY	320	---FGYNSTAV---SQEWSGSAVLNSDINSIOLFVTRYDVTSDNTNTHOKTASATLYLTDDNN	373
Db	463	sdkfandpilkdqgewsagatftsdkgirlyf--dysgkhygkgsittaqvvnvsksd	520
QY	374	GNVSLAQVRNDYIVPEGGGYYQYDQW--KATNKGADNIAMDAHVIEDGNGDRYLVPE	431
Db	521	dtikngvedhktifdgdgkyqnvqgfidegnytsdgnhtlrdphyved-kghkylve	579
QY	432	ASTGLEN-YOGEQDTYNNLWYGGDDAFNIKSLPRILSNDDIKSRATWANAATGILKLNKD	490
Db	580	antctengyqgeeslfnkayyggtnfrfresqk-lqqsakkrdaelangaIgielnd	638
QY	491	EKNPKVAELYSPILASAPWVSEIERPNVVLGNKRYLFAATRLNRGNSDDAMNANYAVG	550
Db	639	ytllkvkm---plitsntvdeIerauvfkmgkwyIftds:gskmtIdg--insn----	689
QY	551	DNVAMGVVADSLGSKYKPLNDGCVLTASVPANWRTATYSYVAVPVAGKDDQVLVTSYM	610
Db	690	-diymIgyvnsnIsgpykplnkglvIqmglpndvtfcyshafvppa-kgnnvvtismy	747
QY	611	TNRNGVAGKGMDSWAPSFLLQINPDNTTTLAKMTNQG	649
Db	748	thrgffedk--katfapsfImnikgnktsvvnksileqg	784
RESULT	9		
ABG21602			
ID	ABG21602	standard; Protein; 1095 AA.	
AC	ABG21602;		
XX			
DT	18-FEB-2002	(first entry)	
XX		Novel human diagnostic protein #21593.	
DE		Human; Chromosome mapping; gene mapping; gene therapy; forensic;	
KW		food supplement; medical imaging; diagnostic; genetic disorder.	
XX			
OS		Homo sapiens.	
XX			
PN	WO200175067-A2.		
PD	11-OCT-2001.		
PF	30-MAR-2001; 2001WO-US08631.		
PR	31-MAR-2000; 2000US-0540217.		
PR	23-AUG-2000; 2000US-0649167.		
PA	(HYSE-) HYSEQ INC.		
PI	Drmanac RT, Liu C, Tang YT;		
PI			
DR	WPI. 2001-639362/73.		
DR	N-PSDB; AAS85789.		
XX			
PT	New isolated polynucleotide and encoded polypeptides, useful in		
PT	diagnostics, forensics, gene mapping, identification of mutations		
PT	responsible for genetic disorders or other traits and to assess		
PT	biodiversity		
PS	Claim 20; SEQ ID No 51961; 103pp; English.		
XX			
CC	The invention relates to isolated polynucleotide (I) and		
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,		
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome		
CC	and gene mapping.		





```

Db 522 -----gegksgtptshlhtkaqlpsphvirhqqvrrqhsqklivgtkaistgkal 573
Qy 106 ASLNTSDSKAAQENTNTAKNDTQKAA PANESSEAKNEPANNVNDSSAANKNDQSSKK- 164
Db 574 rtl-----pkak---vflsppn--ldfkvapsilkprksirednrgsqktv 616
Qy 165 NTTAKLNKDAENVVKKAGI---DPNSLTDDOIKALKNMFSAKAKSG---TQMTYNDQFK 218
Db 617 htegdmmnlkikvkaqvltfttallagatqafakennqkayketygvshtrhdmql 676
Qy 219 IADTLKODGRYVVPFFKASEIKNMPAATTQDAQNTIEPLDWDWSPVQDVRTGOVANK 278
Db 677 ipkq--qnekyvppfdqstikniesa-----kgldwdsplqna--dgtvaey 723
Qy 279 NGYOLVIAMGIP--NONDNHLYLYNKYGDNELSHWKNVGP-----FGYNSTAV---S 328
Db 724 ngyhvvfalagapkdaddtsiymfykvvgdndsdwknagrvfkdskdfandpilkdqt 783
Qy 329 QEWSSGSAVLNSDNIOLFTRVDTSDNNTNHQIKASATLYLDNNGVSLAOVRNDYIVF 388
Db 784 qewsgsatfsgdkirlyft--dysgkhygkslttaqvnvksdtdtkinrvdhktif 841
Qy 389 EGDGYIYQYDQW--KATNKGADNIAMRDADHVEDGNGDRYLVFEASTGLEN-YOGEQDI 445
Db 842 dgdgktyqnvqfidegnytsghntirphyved--kghkyivfeantgtengyqgeesl 900
Qy 446 YNWLNYGGDADANIKSLFRILSDNDIKSRATWANAAIGILKLNKDEKNPKVAELYSPLIS 505
Db 901 fnkayvggntfnfrkesqk-lqgsakkrdaelangelangalgielndytlkkvmk---pllt 956
Qy 506 APWVSDTERPNVVKLGNYKYLFAATRLNRGSDNDAMNANYAVGDNVAMGYVADSLTG 565
Db 957 sntvdelatanvfkmgkwyiftdsrsgskmtidg--lnsn-----diymlygvvsnstg 1009
Qy 566 SYKPLNDSGVVLTSAPVANWRTATYSYAVPVAGKDDQVLTYSYTNNGVAGKGMDSW 625
Db 1010 pykplnkeglvqlmgldpndvtfyshfavpqa--kgnnvitsymtrgfedk--katf 1066
Qy 626 APSFLQLQINPDNTTYTLAKMTNQG 649
Db 1067 apsfllmnikgnktsvkvnsilepg 1090

RESULT 11
ABG04307
ID ABG04307 standard; Protein; 527 AA.
AC ABG04307;
DT 13-FEB-2002 (first entry)
XX Novel human diagnostic protein #4298.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
OS WO200175067-A2.
PN 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR

```

```

DR N-PSDB; AAS68494.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 20; SEQ ID No 34666; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridization probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 527 AA;
SQ

```

```

Query Match 18.4%; Score 758.5; DB 22; Length 527;
Best Local Similarity 37.6%; Pred. No. 6.5e-39;
Matches 194; Conservative 90; Mismatches 181; Indels 51; Gaps 19;

Qy 155 KNDDQSSKK--NPTAKLNKDAENVVKKAGI---DPNSLTDDOIKALKNMFSAKAKSG- 208
Db 37 rednrgsrqktvhtegdmnmnikivkqavltfttallagatqafakennqkaykety 96

Qy 209 --TQMTYNDQFKIADTLIKODGRYVVPFFKASEIKNMPAATTQDAQNTIEPLDWDWSP 266
Db 97 gvshtrhdmqlpkq--qnekyvppfdqstikniesa-----kgldwdsplqna 144

Qy 267 VQDVRTGOVANNYGVOLVIAMGIP--NONDNHLYLYNKYGDNELSHWKNVGP----- 319
Db 145 lqna--dgtvaeyngyhvfalagapkdaddtsiymfykvvgdndsdwknagrvfkdsk 203

Qy 320 FGYNSTAV---SOEWSGSAVLNSDNIOLFTRVDTSDNNTNHQIKASATLYLDNNGV 376
Db 204 fdandpilkdqtqewsgsatfsgdkirlyft--dysgkhygkslttaqvnvksdtdt 261

Qy 377 SLAQVRNDYIVFEGDGYIYQYDQW--KATNKGADNIAMRDADHVEDGNGDRYLVFEAST 434
Db 262 kingvedhktifdgdgktyqnvqfidegnytsghntirphyved--kghkyivfeant 320

Qy 435 GLEN-YOGEQDIYNWLNLYNGGDDAFNFKSLFRILSDNDIKSRATWANAAIGILKLNKDEKN 493
Db 321 gtengyqgeeslfnkayvggntfnfrkesqk-lqgsakkrdaelangelangalgielndytl 379

Qy 494 PKVAELYSPLISAPMVVSDTERPNVVKLGNYKYLFAATRLNRGSDNDAMNANYAVGDNV 553
Db 380 kkvmk---plltentvdeieranvfkmgkwyiftdsrsgskmtidg--lnsn-----di 429

Qy 554 AMVGYVADSLTGSYKPLNDSGVVLTSAPVANWRTATYSYAVPVAGKDDQVLTYSYTNNGV 613
Db 430 ymlgyvsnstltpykplnktglvqlmgldpndvtfyshfavpqa--kgnnvitsymtr 488

Qy 614 NGVAGKGMDSWAPSFLLQINPDNTTYTLAKMTNQG 649
Db 489 gffedk--katfapflmnikgnktsvkvnsilepg 522

```



RESULT 12	
ABG23205	ID
XX	ABG23205 standard; Protein; 789 AA.
XX	AC
XX	ABG23205;
XX	AC
DT	18-FEB-2002 (first entry)
XX	DE
XX	Novel human diagnostic protein #23196.
XX	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder.
KW	XX
XX	OS
XX	Homo sapiens.
XX	XX
PN	WO200175067-A2.
XX	XX
PD	11-OCT-2001.
XX	XX
XX	30-MAR-2001; 2001WO-US08631.
XX	XX
PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	XX
PA	(HYSE-) HYSEQ INC.
XX	XX
PI	Drmanac RT, Liu C, Tang YT;
XX	XX
DR	WPI; 2001-639362/73.
DR	N-PSDB; AAS87392.
XX	XX
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity.
XX	XX
PS	Claim 20; SEQ ID No 53564; 103pp; English.
XX	XX
CC	The invention; relates to isolated polynucleotide (I) and
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC	and gene mapping, and in recombinant production of (II). The
CC	polynucleotides are also used in diagnostics as expressed sequence tags
CC	for identifying expressed genes. (I) is useful in gene therapy techniques
CC	to restore normal activity of (II) or to treat disease states involving
CC	(II). (II) is useful for generating antibodies against it, detecting or
CC	quantitating a polypeptide in tissue, as molecular weight markers and as
CC	a food supplement. (II) and its binding partners are useful in medical
CC	imaging of sites expressing (II). (I) and (II) are useful for treating
CC	disorders involving aberrant protein expression or biological activity.
CC	The polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences, ABG00010-ABG30377 represent novel human
CC	diagnostic amino acid sequences of the invention.
CC	Note: The sequence data for this patent did not appear in the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	XX
SQ	Sequence 789 AA;
	Query Match 18.4%; Score 758.5; DB 22; Length 789;
	Best Local Similarity 37.6%; Pred. No. 1.1e-38;
	Matches 194; Conservative 90; Mismatches 181; Indels 51; Gaps 19;
QY	155 KNDQOQSKK--NTAKLNKDAENVVKKAGI---DPNSLTDDOIKALNKNFSSKAASKG- 208
	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db	299 rednrgsqkvhtegdmnmikivkqatvlfttallagatgafakennqkaykety 358

RESULT	14
ABG21647	
ID	ABG216
XX	
AC	ABG216
XX	
DT	18 - FEE
XX	
DE	Novel

QY 377 SLAQRNDYIVPEGGGYYQTYDOW--KATNKGADNTAMDAHVIEDNGDRYLVEAST 434  
 Db kingvedhktlfdgdgktyqnvqgfidegnytsghnhtlrdphyved-kghkylvfeant 286  
 QY 435 GLEN-YOGEDQIYNWLYNGDDAFNFKSLPRILSNDDIKSRATWANAAGILKLNDEKN 493  
 Db kingvedhktlfdgdgktyqnvqgfidegnytsghnhtlrdphyved-kghkylvfeant 286  
 QY 494 PKVAELSPILSAPVSDERPNVVKLGKNNKYYLFAATRLNRSGNDDAWNANYAVGDNV 553  
 Db kingvedhktlfdgdgktyqnvqgfidegnytsghnhtlrdphyved-kghkylvfeant 286  
 QY 554 AMVGYVADSLTGYKPLNDGSGVLTASVPANWRATYXYAVPVAGKDDOVLVTSWTVNR 613  
 Db kingvedhktlfdgdgktyqnvqgfidegnytsghnhtlrdphyved-kghkylvfeant 286  
 QY 614 NGVAGKGMDSWAPSLLOINPDNTTTLAKMTNQG 649  
 Db kingvedhktlfdgdgktyqnvqgfidegnytsghnhtlrdphyved-kghkylvfeant 286  
 QY 649 gffedk--katfapsflmknkgtsvvksnslleg 488

## RESULT 15

ABG21832  
 ID ABG21832 standard; Protein; 546 AA.

AC ABG21832;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #21823.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 food supplement; medical imaging; diagnostic; genetic disorder.

QX Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Llu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-PSDB; AAS86019.

PT New isolated polynucleotide and encoded polypeptides, useful in  
 diagnostics, forensics, gene mapping, identification of mutations  
 responsible for genetic disorders or other traits and to assess  
 biodiversity

PS Claim 20; SEQ ID No 52191; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
 polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 and gene mapping, and in recombinant production of (II). The  
 polynucleotides are also used in diagnostics as expressed sequence tags  
 for identifying expressed genes. (I) is useful in gene therapy techniques  
 to restore normal activity of (II) or to treat disease states involving  
 (II). (II) is useful for generating antibodies against it, detecting or  
 quantitating a polypeptide in tissue, as molecular weight markers and as  
 a food supplement. (II) and its binding partners are useful in medical  
 imaging of sites expressing (II). (I) and (II) are useful for treating  
 disorders involving aberrant protein expression or biological activity.  
 The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations  
 responsible for genetic disorders or other traits to assess biodiversity  
 and to produce other types of data and products dependent on DNA and  
 amino acid sequences. ABC00010-ABC0377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 546 AA;

Query Match 18.3%; Score 756; DB 22; Length 546;  
 Best Local Similarity 36.9%; Pred. No. 9.7e-39;  
 Matches 195; Conservative 95; Mismatches 189; Indels 50; Gaps 19;  
 QY 141 KNEPAINVNDSSAAKNDQSSKK-NTTAKLNKDAENVVKKAGI---DPNSLTDDQIKAL 196  
 Db 43 kvapsilkprksirednrgsrqktvhtegdmnmikikvqatvltftallagatqaf 102  
 QY 197 NKMNFSKAAKSG---TQMTYNDFOKIADTLIKQDGRYTVPEFKASEIKNMPAATTKDAQ 253  
 Db 103 akennqkayetygvshtrhdmqlpkq--qgnekypvqfdqgstikniesa----- 153  
 QY 254 NTIEPLDVMSWPVQDVRTGVANWNGYQLVIAMGIP-NONDNHIVLLYKNYKGNELSH 312  
 Db 154 ---kgldvwdswplqna-dgtvaeingyhyvvalagapkdaddtslymfqykvgnsls 209  
 QY 313 WKNVGPI-----FCYNSTAV---SQEWSGSAVLNSDNSIQLFYTRVDTSDNNTHQKIA 363  
 Db 210 wkagrvfkdsdkfdandpilkdtqewsgsatftsgdkrlrlyt--dysgkhygkslt 267  
 QY 364 SATLYLDNNGNVSLAQVRNDYIVFEGDGYIYQTYDOW--KATNKGADNTAMDAHVIED 421  
 Db 268 taqvnvksdtklnrvdhkhtifdgdgktyqnvqgfidegnytsghnhtlrdphyved 327  
 QY 422 GNGDRYLVEASTGLEN-YOGEDQIYNWLYNGDDAFNFKSLPRILSNDDIKSRATWANA 480  
 Db 328 -kghkylvfeantgtengyqgeeslfnkayyggntnfrkesqk-lqqsakkrdaelang 385  
 QY 481 AIGILKLNKDEKNPKVAELSPILSAPVSDERPNVVKLGKNNKYYLFAATRLNRSGNDD 540  
 Db 386 algillelnndytlkkvmk---piltstvtdeleeranyfkmngkwyiftdsrsgskmtidg 442  
 QY 541 AMNANYAVGDNVAMVGYVADSLTGYKPLNDGSGVLTASVPANWRATYXYAVPVAGK 600  
 Db 443 --insn-----dlymgyvsnstgpykplnktglvlqmgldpndvftfysheavpqa-k 494  
 QY 601 DDQVLVTSYMTNRNGVAGKGMDSWAPSLLOINPDNTTTLAKMTNQG 649  
 Db 495 gnnvvtstymnrgffedk--katfapsflmknkgtsvvksnslleg 541

Search completed: September 26, 2002, 18:26:15  
 Job time: 154 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2002, 22:12:59 ; Search time 27.35 Seconds  
(without alignments)  
704.636 Million cell updates/sec

Title: US-09-995-587A-1  
 Perfect score: 4129  
 Sequence: -1 MYKSGKNWAVTTLSTAALVF.....GMGTLLSMFGIAEINRRRN 789

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```
Database : Issued_Patents_AA:*
1: /cgn2_6/pdata/2/taa/5A_COMB.pcp:*
2: /cgn2_6/pdata/2/taa/5B_COMB.pcp:*
3: /cgn2_6/pdata/2/taa/6A_COMB.pcp:*
4: /cgn2_6/pdata/2/taa/6B_COMB.pcp:*
5: /cgn2_6/pdata/2/taa/pCTUS_COMB.pcp:*
6: /cgn2_6/pdata/2/taa/backfiles.pcp:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query			ID	Description
		Match	Length	DB		
1	606.5	14.7	455	2	US-08-870-827-3	Sequence 3, Appli
2	281	6.8	415	1	US-08-381-936-2	Sequence 2, Appli
3	281	6.8	415	3	US-08-943-374-2	Sequence 2, Appli
4	279	6.8	578	4	US-09-503-172A-2	Sequence 2, Appli
5	266	6.4	543	1	US-08-362-232-2	Sequence 2, Appli
6	266	6.4	543	1	US-08-814-196-2	Sequence 2, Appli
7	254.5	6.2	1164	4	US-08-923-992A-10	Sequence 10, Appli
8	246.5	6.0	1128	4	US-08-923-992A-6	Sequence 6, Appli
9	242	5.9	984	1	US-08-243-932-2	Sequence 2, Appli
10	242	5.9	984	1	US-08-714-481-2	Sequence 2, Appli
11	242	5.9	984	5	PCT-US95-06111-2	Sequence 2, Appli
12	241	5.8	1164	4	US-08-923-992A-2	Sequence 2, Appli
13	240.5	5.8	1098	4	US-08-923-992A-8	Sequence 8, Appli
14	238	5.8	1104	4	US-08-923-992A-4	Sequence 4, Appli
15	220	5.3	826	1	US-07-638-431-2	Sequence 2, Appli
16	220	5.3	826	5	PCT-US92-00018-2	Sequence 2, Appli
17	213	5.2	1565	6	5352450-2	Sequence 2, Appli
18	194	4.7	1912	1	US-08-409-995-4	Sequence 4, Appli
19	194	4.7	1912	3	US-08-688-467-4	Sequence 4, Appli
20	194	4.7	2411	4	US-09-268-347-36	Sequence 36, Appli
21	192	4.7	2353	4	US-09-377-155-33	Sequence 33, Appli
22	192	4.7	2353	4	US-08-913-942-4	Sequence 4, Appli
23	192	4.7	2353	4	US-09-669-974-33	Sequence 33, Appli
24	191.5	4.6	1561	3	US-08-894-017-23	Sequence 23, Appli
25	189	4.6	2354	4	US-09-268-347-47	Sequence 47, Appli
26	184	4.5	1430	3	US-09-008-172-2	Sequence 2, Appli
27	184	4.5	1430	4	US-09-210-361-6	Sequence 6, Appli

Query Match	14.7%	Score 606.5;	DB 2;	Length 455;
Best Local Similarity	36.2%	pred No	4	20-26.

QY 229 RYTPFFKASEIKNMPAAATTKDAQTNTIEPLDYWDSWPQDVRTGQVANWNGYQLVIAMM 288  
:: || || || || || : | | : || || || || : | || :: || : | :

## ALIGNMENTS

```

1
RESULT
US-08-870-827-3
; Sequence 3, Application US/08870827
; Patent No. 5962297
; GENERAL INFORMATION:
; APPLICANT: Tsusaki et al.
; TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
; - TITLE OF INVENTION: ACTIVITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,827
; FILING DATE: 06-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 170,630/1996
; FILING DATE: 10-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TSUSAKI-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 455 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-870-827-3

```

MOLECULE T  
US-08-870-827-3

Db 31 QKVPQFNASAKNIDSAGYDKSGNLID-LDWDSWPLQNA-DGTAANYHGHIYSALA 88  
 QY 289 GIP-NONDNHIIYLLYKNGDNLHSHKNGVPIFYNSTAV-----SQWSSGSAVLN 338  
 Db 89 GDRKNSDDTPLHLFYQKVGDTSDSKNAGRVPEDMDKFVPNDPKLYKQTEWSSSATIT 148  
 QY 339 SDNSIQLYTRVDTSON-----NTHOKIASATLYLTD-NNGNVSIAQVRNDYIVFE-G 390  
 Db 149 KGOVRLEVT--DYSNPDGGTGAGNQIISTAQVNLSDPDAATLKVDGVSDBKSVFDGG 206  
 QY 391 DGYYQYDOWKATNK--CADNIAMRDAHVIEDGNGDRYLVPFASTG-LENYOGEDQIYN 447  
 Db 207 DGTVYQIQOIFDEGKWISGDNHTLRDPHYVED-KGHKYLVEATGTTDYGQGGQSPNN 265  
 QY 448 WLVYGGDDAFNIKSLFRILSNDNDIKSRATWANAAGILKLNDEKNPKVAELYSPLISAP 507  
 Db 266 KAYYGGSDVFEQNEKNKLLQSPK-KQIASLANGALGIVELADDY---TVKSVMKPLVASN 321  
 QY 508 MYSDETERPNVYKLNKYYLFAATRLNRGSNDAMNANYAVGD-NVAMVGYVADSLTGS 566  
 Db 322 TVADEVERANIFKMNKWLFTDSRGSKWTSD-----GINKDYVYMLPGGDSLNGP 373  
 QY 567 YKPLNDGVLVTASYPANWRTATSYAVPVAGKDDQVLVTSYMTNRN----- 614  
 Db 374 HNPINETGLVNLMDPLADLTHYSHGIP-HPEGNNVVLTSYMTNRGTFYPEHSHLRDK 432  
 QY 615 -GVAGKGHMDST 624  
 Db 433 LGVNIKGS DTS 443

## RESULT 2

US-08-381-936-2  
 ; Sequence 2, Application US/08381936  
 ; Patent No. 5792923

## GENERAL INFORMATION:

APPLICANT: ROBER, Manuela  
 APPLICANT: GEIER, Gebhardt  
 APPLICANT: GEIDER, Klaus  
 APPLICANT: WILLMITZER, Lothar  
 TITLE OF INVENTION: DNA sequences which lead to the  
 TITLE OF INVENTION: formation of polyfructans (levans), plasmids containing  
 TITLE OF INVENTION: these sequences as well as a process for preparing  
 NUMBER OF SEQUENCES: 2

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen  
 STREET: 1180 Avenue of the Americas  
 CITY: New York  
 STATE: NY  
 COUNTRY: US

ZIP: 10036-8403

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/381.936

FILING DATE: 09-FEB-1995

## CLASSIFICATION DATA:

PRIOR APPLICATION DATA:

FILING DATE: 09-AUG-1993

## APPLICATION DATA:

APPLICATION NUMBER: DE P42270618

FILING DATE: 08-DEC-1992

## ATTORNEY/AGENT INFORMATION:

NAME: Meilman, Edward A.

REGISTRATION NUMBER: 24,735

REFERENCE/DOCKET NUMBER: P/951-108

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 382-0700

TELEFAX: (212) 382-0888  
 TELEX: 236925  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 415 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-381-936-2

## Query Match

Best Local Similarity 6.8%; Score 281; DB 1; Length 415;

Matches 109; Conservative 73; Mismatches 137; Indels 106; Gaps 22;

QY 257 EPIDVDSWPQDVRTGQVANNNGYQIVIAMMGIPN-----QNDNHIYLL-----YNYKG 306  
 Db 40 EEVPIINDMPLRDF-DGEIISVNGWCIIFLTADRNTDNPQFODENGNDYDITRDWEDRHG 98  
 QY 307 DNELSHW-KNVGP--IFGYNSTA-----VSQWSSGSAV-LNSDNSTQLFVTVRVDTSNNT 357  
 Db 99 RARICWYSRTGKDWIFGGRVMAEGVAPTRWAGTPILLNDRGDDIDLYTCV-----T 152  
 QY 358 NHOKIASATLYLTDNNGNYSIAQVRNDYIVFEEDGYYQYTDQ---WKATNKGADNIAMR 414  
 Db 153 PGATIAKVRGKIVTSDQSVSLGEGFOQVTSLSADGTIYOTEEQNAPFN-----FR 202  
 QY 415 D-AHVIEDGNGDRYLVPFASTGLENYQGEDQIY-----NWLNYGGDDAFNIKSLFR 464  
 Db 203 DPSPFIDRNDGKLYMLFEGN--VAGPRGSHETQAEEMGNVPPGYEDVGG----- 249  
 QY 465 ILSNDDIKSRATWANAAGILKLNDEKNPKVAELYSPLISAPVSDIERPNVYKLNK 524  
 Db 250 -----AKYQACGVG-LAVAKDLGSE-WQLPLPLITAVGVNDQTERPHEVFDGK 297  
 QY 525 YLFAATRLNRGSNDAMNANYAVGDVNY-----AMGVVYVADSLTGSYKPLNDGVLVTAS 580  
 Db 298 YLFT-----ISHKYTFADNLGTGPDGVYGVPSDKLTGCTPYTPMNSSGLVL--G 342  
 QY 581 VPANWRTATSYAVPVAGKDDQVLVTSYMTNRNGVAGKGM-----STWAPSLQLQINPD 636  
 Db 343 NPSSQPEQTSYHYNP-----NGLVTSFI---DSVPMKGRDYRIGGTGTEAPTAKILLKGD 393  
 QY 637 NTTVV 641  
 Db 394 RSFIV 398

## RESULT 3

US-08-943-374-2

; Sequence 2, Application US/08943374

; Patent No. 6028249

## GENERAL INFORMATION:

APPLICANT: ROBER, Manuela  
 APPLICANT: GEIER, Gebhardt  
 APPLICANT: GEIDER, Klaus  
 APPLICANT: WILLMITZER, Lothar  
 TITLE OF INVENTION: DNA sequences which lead to the  
 TITLE OF INVENTION: formation of polyfructans (levans), plasmids containing  
 TITLE OF INVENTION: these sequences as well as a process for preparing  
 NUMBER OF SEQUENCES: 2

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen  
 STREET: 1180 Avenue of the Americas  
 CITY: New York  
 STATE: NY  
 COUNTRY: US

ZIP: 10036-8403

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

```

: APPLICANT: TONOUZUKA, Takashi
: APPLICANT: SAKANO, Yoshiyuki
: TITLE OF INVENTION: BETA-FRUCTOFURANOSIDASE GENE
: FILE REFERENCE: 10749-0001-0
: CURRENT APPLICATION NUMBER: US/09/503,172A
: CURRENT FILING DATE: 2000-02-14
: PRIOR APPLICATION NUMBER: JP 160416/1999
: PRIOR FILING DATE: 1999-06-08
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 2
: LENGTH: 578
: TYPE: PRT
: ORGANISM: Arthrobacter sp.
: US-09-503-172A-2

```

```

Query Match      6.8%   Score 279; DB 4;   Length 578;
Best Local Similarity 23.4%; Pred. No. 3.3e-12;
Matches 139; Conservative 83; Mismatches 207; Indels 164; Gaps 28;

QY 131 RAPANESSEAKN--EPAVNVNDSAAKNDDQSSKNNTTAKLNKQENNVVKAGIDPNSL 188
      ||||| | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
db 31 AAPATAOGLDGPETHTQAYAVEDD-----FTAK----- 63

```

```

Query Match          6.8%; Score 281; DB 3; Length 415;
Best Local Similarity 25.6%; Pred. No 1.4e-12;
Matches 109; Conservative 73; Mismatches 137; Indels 106; Gaps 22;

QY 257 EPLDVMDSKPVQDYRTGVQVANKNGYQLVIAMGIPN-----QNDNHIYLL-----YNKYG 306
Db      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
40 ERVFTWDTMPLRDF-DGEIISVNGWCIFTLTADRTDNFQFDENGNYDITROWEDRHG 98
QY 307 ONELSHW-KNYGP--IFGYNSTA-----VSEWGSQAV-LNSDNSIQLFYTRVDTSDNNT 357
Db      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
99 RARICYWYSRTGKWDFIEGGRVMAQSGVAPTPREWAGTPILLNDRGDIDLITYTCV-----T 152
QY 358 NHQIKASATLYLTNNNGNSVLAQVRNDYIYFEGDGYVYYQYDQ---WKATNKGADNIAKR 414
Db      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
153 PGATIAKVRGKIVSDQSLSGFGQQVTSLSADGTIYQTEQNAFWN-----FR 202
QY 415 D-AHVIEDNGDRYLVEASTGLENTQGEDQIY-----NWLNVGGDDAFNKSLEPR 464
Db      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
203 DPSPFIDRNDGKLYMLEPENG--VAGPRGSHETIQAENGWVPPCYEDVGG-----249
QY 465 ILSNDDIKSRATWANAAGITLKNLKEKNPKVAELYSPLISAPWVSEIERPNVYKLGNK 524
Db      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
250 -----AKTQACGVG--LAVAKDLGSE-WQILPLPLITAVGVNDQTERPHFVFQDGK 297
QY 525 YLFLAATRLNLGRNSDDAWMNANTAVGDNY-----AMVGYVADSLTGSYKPLNDOSGVVLTS 580
Db      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
298 YLFLT-----ISHKYTFADNLTPGDGYGVFVSDKLTGPYTPMNSGLVL--G 342
QY 581 VPANRMTATSYXVAVPVACKDDQVLVTSYWTNRNGVAGKGM-----STWASFLQINPD 636
Db      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
343 NPSSQFQYIYSHVMP-----NGLVTSFI-----DSVPWPKGKYRIGGTEAPTVKILKGD 393
QY 637 NTTTV 641
Db      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
394 RSFIV 398

```

RESULT 4

US-09-503-172A-2

; Sequence 2, Application US/09503172A

; Patent NO. 6284510

; GENERAL INFORMATION:

; APPLICANT: ITO, Tetsuya

; APPLICANT: FUJITA, Koki

; APPLICANT: HARA, Kozo

```

RESULT      5
US-08-362-232-2
; Sequence 2, Application US/08362232
; Patent No. 5641667
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION:  Fructosyltransferase Enzyme, Method
; TITLE OF INVENTION:  For its Production and DNA Encoding the Enzyme.
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann & Baron

```

```

; STREET: 350 Jericho Turnpike
; CITY: Jericho
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11758
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk, 3.5 inch, 1.44 Mb
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 6.0 for DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,232
; FILING DATE: 22-December-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CU 125/93
; FILING DATE: 23-December-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Baron, Ronald J.
; REGISTRATION NUMBER: 29,281
; REFERENCE/DOCKET NUMBER: 294-29
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 822-3550
; TELEFAX: (516) 822-3582
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-08-362-232-2

```

Query Match 6.4%; Score 266; DB 1; Length 543;  
 Best Local Similarity 26.9%; Pred. No. 2.6e-11;  
 Matches 141; Conservative 63; Mismatches 171; Indels 150; Gaps 35;

```

QY 209 TOMTY---NDFQ---KIADPL-IKQGRYTPVPEKASEIKNMPAATK---DAQTNTIEP 258
DB 33 TOQAYDPQSDFTARWTRADALQIKAHSDATV---AAGQNSLPAQLTMNIPADFPVNP 88
QY 259 -LDVMSWPVQDVRTQGVANWNGYQLVIAMMGIPNO-----NDNHVY-----LLYNYKGDNE 309
DB 89 DVWVMDTWLIDKHADQF-SYNGWEVIFCLTADPNAGYGFDDRRHVHARIGFFYRAGIPA 147
QY 310 LSH-----WKNVGPPIGYNSTAV-----SQWSSGSAVLNS--DNSIOLFYTFRV--- 350
DB 148 SRPVPNGGWTYGGHLPFGDASQVYAGQYTNQAEWSSGSRLLMQIHNVTVSFYTDVAFN 207
QY 351 -DTSNN-TNHOKIASATLYLTDNNGVNSLAQVRNDY---IVFEG-----DGYYVQ 396
DB 208 ROANANNITPPOAITQT-----LGRHADFNHNVWFTGTFTHTPLLPQDGVLYQ 256
QY 397 TYDQWKATNKGADN--IAMRDAHVIEDGN--GDRYLVEAST-----GLENYQGEDOIYNW 448
DB 257 -----NGAQNEFFNRDPFTEDPKHPGVNVMVEGNTAGORGVANCTEAD----- 302
QY 449 LNYGGDAFNKSLFRILSNDDIKSRATWANAAGILKLNKDEKNPKVAELYSPLISAPM 508
DB 303 LGFRNDP-NAETLOEVLN-----SGAYYQKANTIG-LAIATDSTLSK-WKFLSPLISANC 354
QY 509 VSEITERPNVVLGNKYYLFAATRLNRGSDNDANWANYAVG-----DNVAMGVVADSLT 564
DB 355 VNDQTERPQVYLHNGKYYIFISH-----RTFAAGVDGPDGV---YGEVGDGIR 401
QY 565 GSKPLNDSGVVLVATASVPANWRTA-----TYSYAVVPVAGKDDQVLVT 607
DB 402 SDFQPMN-YGSGLTGNGMPTDNLNTAAGTDFDPSDQNPRAFOSSHYVMPGG-----LVE 454
QY 608 SY---MTNRNGVAGKMDSTWAPSFLLQINPDNTTTVLAKTNQOG 649
DB 455 SFDITVENRRG-----GTLAPTVRVRI-AONASAVDLRYNGG 491

```

## RESULT 6

```

US-08-814-196-2
; Sequence 2, Application US/08814196
; Patent No. 5731173
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Fructosyltransferase Enzyme, Method
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann & Baron
; STREET: 350 Jericho Turnpike
; CITY: Jericho
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11758
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk, 3.5 inch, 1.44 Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 6.0 for DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/814,196
; FILING DATE: 10-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,232
; FILING DATE: 22-December-1994
; APPLICATION NUMBER: CU 125/93
; FILING DATE: 23-December-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Baron, Ronald J.
; REGISTRATION NUMBER: 29,281
; REFERENCE/DOCKET NUMBER: 294-29
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 822-3550
; TELEFAX: (516) 822-3582
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-08-814-196-2

```

Query Match 6.4%; Score 266; DB 1; Length 543;  
 Best Local Similarity 26.9%; Pred. No. 2.6e-11;  
 Matches 141; Conservative 63; Mismatches 171; Indels 150; Gaps 35;

```

QY 209 TOMTY---NDFQ---KIADPL-IKQGRYTPVPEKASEIKNMPAATK---DAQTNTIEP 258
DB 33 TOQAYDPQSDFTARWTRADALQIKAHSDATV---AAGQNSLPAQLTMNIPADFPVNP 88
QY 259 -LDVMSWPVQDVRTQGVANWNGYQLVIAMMGIPNO-----NDNHVY-----LLYNYKGDNE 309
DB 89 DVWVMDTWLIDKHADQF-SYNGWEVIFCLTADPNAGYGFDDRRHVHARIGFFYRAGIPA 147
QY 310 LSH-----WKNVGPPIGYNSTAV-----SQWSSGSAVLNS--DNSIOLFYTFRV--- 350
DB 148 SRPVPNGGWTYGGHLPFGDASQVYAGQYTNQAEWSSGSRLLMQIHNVTVSFYTDVAFN 207
QY 351 -DTSNN-TNHOKIASATLYLTDNNGVNSLAQVRNDY---IVFEG-----DGYYVQ 396
DB 208 ROANANNITPPOAITQT-----LGRHADFNHNVWFTGTFTHTPLLPQDGVLYQ 256
QY 397 TYDQWKATNKGADN--IAMRDAHVIEDGN--GDRYLVEAST-----GLENYQGEDOIYNW 448
DB 257 -----NGAQNEFFNRDPFTEDPKHPGVNVMVEGNTAGORGVANCTEAD----- 302

```



316	266	281	433	337	490	397	534	448	584	494	632	548	692	604	742	658	783	705	842
Db	QY	QY	Db	QY	Db	QY	Db	QY	Db	QY	Db	QY	Db	QY	Db	QY	Db	QY	Db
<div> <div>RESULT 8</div> <div>US-08-923-</div> <div>Sequence</div> <div>Patent No</div> <div>GENERAL</div> <div>APPLIC</div> <div>APPLIC</div> <div>TITLE</div> <div>TITLE</div> <div>CORRES</div> <div>CORRES</div> <div>ADDRESS</div> <div>STREET</div> <div>CITY</div> </div>																			

[illegible]



RESULT 10  
US-08-714-481-2  
; Sequence 2, Application US/08714481  
; Patent No. 5766606  
; GENERAL INFORMATION:  
; APPLICANT: Brady, L. Jeannine  
; TITLE OF INVENTION: Cloning of No. 5766606-IgA FC Binding Forms of  
; TITLE OF INVENTION: the group B Streptococcal Beta Antigens  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saliwanchik & Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville

Query Match	5.9%	Score 242;	DB 1;	Length 984;
Best Local Similarity	20.7%;	Pred. No. 3.5e-09;		
Matches 176;	Conservative 112;	Mismatches 325;	Indels 238;	Gaps
QY	6	KNWAVVLTSTAALVFGATTVNASADTNIENDSST-----VOVTTGCDNDIAVKSVTGLGS	59	
DB	5	KDDSVKTTTEVAAPKY-----PSMAQTQDGNNSSSSELTTKMBIPITDITKKAAVEPVEKTA	59	
QY	60	GOVSAASDTRTISA-----NANSASSAANTQNSQVASSAAITSTSSAASI-----	108	
DB	60	GETS-ATDTGKREQLQOQWKNLKNVDVNTILSHEQKNEKTKIDETNDSALLELENOF	118	
QY	109	NNT-----DSKAAQOENTYAKNDTOTKAAAPANSESSEKAPVAVNVNDSSAA	154	
DB	119	NETNRLHLIKOHEVEKDKAKAQO---KTLKQST-KVDLSNIDKE-----LNHKQSQVE	169	
QY	155	KNDDQSSKKNNTAKLNKDAENVVKAGIDPNSLTDDQIKALNKNMFSSKAAKSGTQMTYN	214	
DB	170	KMAQGGITINEDKSMCLKKTETRKQA-----QOADKKEADAEVKVQLEEEAHSKLQVVE	224	
QY	215	DFQKIADTLTIKODGRYTVPPFKASE-----IKNPAATTQKAO---TNTIEPLDWD	263	
DB	225	DFRK-----KFTSEQVTPKRVKRDLANENNNOQKIELTVSPENITVYE	269	
QY	264	SWPQDQVRGQVANNQYOLVIAMGIPNQNHHYL-----LYNKYGDNELSHKNVGP	319	
DB	270	G---EDVK-----FTVTAKSDSKTTLDFSLLTKY-----	296	
QY	320	FGYNSTAVSOWSGSVALNSDINSIQLEYTRYDTSNNTHQKIASATLYLTDDNGNVSLA	379	
DB	297	---NPVSDRISTNYKTTN-----HKIAEITIKNLKLNESQTVTLKAKDDSGNY---	344	
QY	380	QVRNDYIVFEGDGYQYTDQWKATNKGADNIAMDRAHVE-----DGNCDRYLVE	430	
DB	345	-VEKTFITV-----QKKEEKQVPKTPQEKDSKTEKVPQEPKSNCKNQLQELIK	393	
QY	431	EASTGLENYQGEDQIYNWLNLYGDDAFNIKSLERTLSNDDIKS-----RATWAN---	479	
DB	394	SAQOELKLEK-----KAIKELMEQPEIPSNPEYGIQKSTWESQKEP	434	
QY	480	---AAIGILKLNKDEKNPKVAELYSPLISAPMVSEIERP-----NWKLGKNKYLF	530	



;; TITLE OF INVENTION: NO. 6280738-IgA Fc Binding Forms of the Group B  
;; NUMBER OF SEQUENCES: 34  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
;; STREET: 1100 New York Avenue, N.W., Suite 600  
;; CITY: Washington  
;; STATE: D.C.  
;; COUNTRY: USA  
;; ZIP: 20005  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/923,992A  
;; FILING DATE: 05-SEP-1997  
;; CLASSIFICATION: 536  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 60/024,707  
;; FILING DATE: 06-SEP-1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Esmond, Robert W.  
;; REGISTRATION NUMBER: 32,893  
;; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (202) 371-2600  
;; TELEFAX: (202) 371-2540  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1164 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-923-992A-2

Query Match 5.88; Score 241; DB 4; Length 1164;  
Best Local Similarity 20.08; Pred. No. 5.3e-09;  
Matches 191; Conservative 130; Mismatches 347; Indels 288; Gaps 40;

QY 10 VVTLSTALVFGA-----TTVNAS-----ADTNINNDST-----VQV 43  
DB 21 VASVAVASLPMGSAHASSELVKDVSVKTEVAAPYPSMAQTDQGNSSSELETTKMEI 80  
QY 44 TTGNDIAVKSVTILGSGVSAASDITRTSA-----NANSASSAANTQNSQVASSAAIT 99  
DB 81 PTTDIKKAPEVPEKTAGEYS-ATDTGKREKQLQKQWNNLNKNDVDNTILSHEQKNEFKTKI 139  
QY 100 SSTSSAASL-----NNT-----DSKAAQENTNTAKNDTQ-----129  
DB 140 DETNDSALLELENQFNETNRLHLTKHQBEEVEKDKKAKQ--KTLKQSDTKVDLSNDIK 197  
QY 130 ---KAAPANESSAK---NEPAVNVNDSSAAND---QSSKK-----164  
DB 198 LNHQKSOVEKMAEQGITNED---KDSMLKIEDIRKQAQADKEDAEVKVREELGKL 253  
QY 165 --NTTAKLNKDNVVK---AGID---PNSLTDQIKALNKNFSAKAKSGTQM 211  
DB 254 FSTKAGLDQEQEHVKTKETSSEENTQKVDEHYANSLQNLAKLSLELDKATTTNEQATQV 313  
QY 212 TYNDF-----QKTIAD---TLIKODGRYTVPFKA-----SETKNMPAATTKDAQTNTI 256  
DB 314 K-NQPLENAQKLKEIQLIKET---NVKLYRAMSESLEQVEKELKHNSLEANLDAVASK 369  
QY 257 EPLDWDS-----  
DB 370 ELVREYEGKLNQSKNLPKLEBEAHSKLVQVVEDFRKKFKTSEQVTPKKRVKRLAAN 429  
QY 278 WNGYQ---LVITAMGIPNDNHHYLLYKNYGNELSHWKNVGPITFCYNSTAVSQBWSGS 334  
DB 430 ENNOQKIELTVPENIVYGEDVKFTVTAKSQSKTT--LDFFSLLTKYNPVSVDRISTN 487

QY 335 AVLNSDNSQLFTRVDTSDNNTNHQKIASATLYLTDNNGVSLAQVRNDYIVFEGDYY 394  
DB 488 YKNTDND-----HKIAETITKLNKLNESQTVTLRAKDDSGNV-----VEKFTTIV-----533  
QY 395 YQTYDQWKATNGKADNIAMROAHVTE-----DGNDRYLVFEASTGLENYQGEDQI 445  
DB 534 -----QKREKQVPTPEQKSKTEEKVPQEPKSNKNDKQELIKSAQOELEKLE-----583  
QY 446 YNNLNYGGDAFNKLSFRILSNDDIKS-----RATWAN-----AAIGILKLNKDE 491  
DB 584 -----KAIKELMEQPEIPSNPEYGIQKSIWESQKEPQEAITSKKIIGDS 629  
QY 492 KNPKVAELYSPLISAPMVSDIERP-----NVVKLGNKYLLPAATLNRGSSNDADWMA 545  
DB 630 SSKYVTEHYFNKYKSDFWYQLHQAEMELTRKVVQVYMNKYPDNAEIKKIFESDMKRTKD 689  
QY 546 NYAVCDNVAMGVYVADSLTGSYKPL-----NDSGVVLTASVPANWRATYTYAVPVAGKD 601  
DB 690 NYGSLNDALKGFEKYFLTPFNKIKQIVDDLDKKVQDQPA-----PIPENSEM 739  
QY 602 DQ-----VLVTSYMTNRNGVAGKGMDSWAPSFLLQINPDNTTVLAKMTNOGDWIDD 655  
DB 740 DQAKERAKIAVSKYMS-----KVLGCVHQ-----HLQKKNSKIV-----DLFKEL 780  
QY 656 SSENLMIGDLSAALPGERDKPVDWDLIG-----YGLK---PHDPATPNDPPTP 702  
DB 781 EAIKQGTIFDIDNAKTEVEIDNLVH-DAFSKMNATVAKFQKLENTTETPTDKRIPELP 839  
QY 703 TTPETPTPTPTKPTPENPGTROTPTNPTNPTPEPLTPETPKQPTQTNNRLPQT 758  
DB 840 QAPDTPQADPTPHVPESPKAPEAPRVPEPKTPEAPHVPEPKAPEAPRVPEPKT 895

RESULT 13  
US-08-923-992A-8  
; Sequence 8, Application US/08923992A  
; Patent No. 6280738  
; GENERAL INFORMATION:  
; APPLICANT: Tai, Joseph Y.  
; APPLICANT: Blake, Milan S.  
; TITLE OF INVENTION: NO. 6280738-IgA Fc Binding Forms of the Group B  
; TITLE OF INVENTION: Streptococcal Beta Antigens  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Avenue, N.W., Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/923,992A  
; FILING DATE: 05-SEP-1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/024,707  
; FILING DATE: 06-SEP-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Esmond, Robert W.  
; REGISTRATION NUMBER: 32,893  
; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1098 amino acids

```

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-923-992A-8

```

Query Match	5.8%;	Score	240.5;	DB	4;	Length	1098;
Best Local Similarity	19.2%;	Pred.	No.	5.3e-09;			
Matches	179;	Conservative	132;	Mismatches	360;	Indels	259;
Gaps	34;						
QY	6	KNWAVVLTSTAALVFCATVNASADTNTENDSST-----VQVTTGDNDAIVASVTLGS	59				
Db	11	KDQSVKTTVAAPY-----PSMAQTDOGNSSSSSELETTTMEIPTTDIKAVEPVEKTA	65				
QY	60	GOVSAAAS-----DPTTTSANANASSAANTQNSNOVASSAITS	100				
Db	66	GETSATHTGKREKQLQWKNLKNVDVNTILSHQKNEFKTIDETNDSALLLENOFN	125				
QY	101	STSSAASLN-----NTDSKAAQEWI-----NTAKNDPTQKAAPANESSEAKNEP	144				
Db	126	ETNRLLLHTHQHEVEKKAKAQKTLKQSDTKVDLSIDRELNHOKSQEAGITNEDKDSM	185				
QY	145	AVNVNDSSAAKNDQQSSKK-----NTTAKLNKDAENVVK-----	180				
Db	186	LKKIED-----IRKQAQDPDKREDAEVKVEELGKLFSTKAGLQOEIQEHVKKETSSEENT	242				
QY	181	AGID-----PNSLTDDQIKALKNFSSKAAGSGTQMYNDF-----OKIAD---TLIKQGRYT	231				
Db	243	QKVEHYANSLQNAQSKSLELDKATTTNEQATQVK-NQFLENQAQKLKEIQPLIKET--N	298				
QY	232	VPPFKA-----SEIKNPAATTKDAQTNTIEPLDVWDS-----	264				
Db	299	VKLYKAMESSELEQVEKELKHNSLEANLDAVAKSEIVREYEGKLNOSKNLPELKQLEEA	358				
QY	265	-----WPQDVQVGOVANNNGYQ---LVIAMMGIPNQNDNHLYL	300				
Db	359	HSKLLQVVEPRFKFTTSEQVTQPKKLRKRLDAENANNQOIKELTSPENITVVEGEDYKF	418				
QY	301	LYNKYGNELSHKNVGPFIIFYNSTAVSQWGSASVLNSDNSITQLFYTVDTSNDNTNHQ	360				
Db	419	TVTAKSSDKTT--LDPSDLLTKYNPSVSDRISTNYKTNTDN-----HKTAETIKNLKN	471				
QY	361	KIASATLYLTDNNGNLSLAQVRNDYIVFEGDGYGYQTYDQWKATNGADNIAHMDAHVIE	420				
Db	472	ESQTVTLKANDSGNV-----VEKTFITV-----QKKEEQVQPKTPBQKDSKTEE	517				
QY	421	-----DGNQDRYLVEASTGLNYESQGEDQIYNWLNLYGGDDAFNTKLSFRILSNDDI	471				
Db	518	KVQPEPKSNDKNLOELIKSAQOEKLE-----KAKELMEQPEI	558				
QY	472	KS-----RATWAN-----AATGILKLNKDEKNPKVAELYSPLISAPVSDIEERP-	516				
Db	559	PSNPEYGIQKSTWESQKEPIQIAITSFVKIIGDSSKYTYTEHYFNKYKSHFEMNYQLHAQM	618				
QY	517	-----NVWKLGNKYLYLAATRLNRGNSDDAWNNANYAGVDNVAWGVYVADSLTGSYKPL-	570				
Db	619	EMLTRKQWQYMKYPDNAEIKKIPESDMKRTKEDNYGSLENDALGKFEKYLFPFNKIK	678				
QY	571	---NDSGVVLTASVANNRTATYSYAYVPVAGKDDQ-----VLVTSYMTNRNGVAGKGM	621				
Db	679	QIVDDFDKKVEQDQA-----PIPENEMDQAKERAKIAVSXKMS-----KVL	721				
QY	622	DSTWAPSFLLQINPDNTTVTLAKMTNQGDDTWDDSSENLMDIGLDSAAALPGERDKPDVW	681				
Db	722	DGVHQ-----HLOKNHSHKIV-----DLFKELEAIKQOTIFDIDNAKTEVELDNLVH-	768				
QY	682	DLIG-----YGLX-----PHDPATPNDEPTTTPETPTETPNTPKTPTKPNPQTPQT	728				
Db	769	DAFSKKNATVAKFQKGLTNTPTDTPKPIELPQAPDTPQAPDTHVHPQSKAPEAPRV	838				
QY	729	PNTPTNPEIPLTPTPKQPEPQTNNRLPQT	758				
Db	829	PESPKTPEAPHVPSPKAPEAPRVPSPKT	858				

```
Db 294 KEIQLIKET---NVKLYKAMSESLQVEKELKHSEANLQDLVAKSEIVREYEGKLNQ 350
QY 265 -----WPQDVRTQGVANWNGYQ---LVIA 286
Db 351 SKNLPKOLEBEASHKLVQVHEFRKFKTSEQVTPKRVKRDIAAENNOCKIELTVS 410
QY 287 MMGIPNQNHHLYLLYNKYGNELSHKNGVPIFYNSTAVSQEWGSAVLNDSNQLF 346
Db 411 PENITVYEGEDVKFTVAKSKTT--LDFSLLTKYNSVSDRISTNYKTNTON-----463
QY 347 YTRVTSNNTNHQIKASATLYLDNNGNVSIAQVRNDYIVF---EGDGYIYQYDQWKA 403
Db 464 HKIAETITKLNKLNQSOTVTLKAKDGSNV---VEKFTTITVQKKEQVKTPEQ--K 517
QY 404 TNKGADNIAARDHAHVIEDGNGDRVLVPEASTGLENYQGEDQIYNWLNYYGGDDAFNIKSLF 463
Db 518 HSKTPONVQBPKS--NDKNQOLQELIKSAQOEKLE-----RAIK 556
QY 464 RILSNDKIS-----RATWAN-----AAIGILKLNKDEKNPKVAELY-----500
Db 557 ELMEQPEIPSNPEYGIQKSIWESQKEPIQEAITSFNKLIIGDSSKYYTEHYFNKYKSHFM 616
QY 501 SPLISAPWVSDEIERPNVYKLGKYYLFAATRLNRGSDNDAMNANYAVGDNVAWGYVA 560
Db 617 NYQLHAQM---EILTRVQVMKYPDPAEIKKIFESDMKRTKEDNYGSLDALKGYPE 673
QY 561 DSLRGSYKPL---NDSGVVLTASVPANWRTATSYAVVPVAGKDDQ-----VLVTSYM 610
Db 674 KYELTPENKIKQIVDDLKVEQDQPA-----PIPENSEMDOAKEKAKIAVSKYM 723
QY 611 TNRNGVAGKMDSTWAPSFLLQNPONTTTLVAKMTNQGDWIMDDSENLDIGDLSAA 670
Db 724 S-----KVLGVHQ-----HLQKKNHSHKIV-----DLFKELEAIKQOTIFDIDNAK 764
QY 671 LPERDKPVDWDLIG-----YGLK---PHDPATNDPETPTTPTETPTNTPKTP 717
Db 765 TEVEIDNLVH-DAFSKNWATVAKFQKLENTPTPTPRIPELQAPDTPQAPDTPHPV 823
QY 718 KTPENGTQTPNTPTPEIPTPKPTPKPTQPTNNRLPOTGN 760
Db 824 ESPKAPAPRVPSNPPTPEAPHVPSKAPPEP---RVPESPN 863
```

## RESULT 15

US-07-638-431-2  
; Sequence 2, Application US/07638431  
; Patent No. 5198535

## GENERAL INFORMATION:

; APPLICANT: Hofman, Stephen L.  
; APPLICANT: Charoenvit, Yupin  
; APPLICANT: Hedstrom, Richard  
; APPLICANT: Khumsmith, Srisin  
; APPLICANT: Rogers IV, William O.  
; TITLE OF INVENTION: Protective malaria sporozoite surface protein  
; TITLE OF INVENTION: Immunogen and gene  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: A. David Spevack  
; STREET: NMRDC Building 1 T-12 National Naval  
; STREET: Medical Center  
; CITY: Bethesda  
; STATE: MD  
; COUNTRY: USA  
; ZIP: 20814-5044

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.24

; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/638.431  
; FILING DATE: 19910110

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:  
; NAME: Spevack, Aviom D.  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 295-6759  
; TELEFAX: (301) 295-4033  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 826 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-638-431-2

Query Match 5.3%; Score 220; DB 1; Length 826;  
Best local similarity 20.7%; Pred. No. 11e-07;  
Matches 87; Conservative 46; Mismatches 132; Indels 156; Gaps 17;

```
QY 415 DAHVIEDGNGDRVLVPEASTGLENYQGE-----DQIYNWLNYYGGDDAFNIKSLFRILSND 469
Db 43 DIHILDGSG-----SIGYSNWKAAHVIPMLNTLVNLSNDEINVSFLTFSTNSRE 94
QY 470 DIKSRATWANAAGIILKLNKDEKNPKVAEL---YSP-----LISAPWVSDEIERP 516
Db 95 LIKLR-----GYGSTKSDSLRFLAHLQNNYSPNGWNTLTSALLVVDTLINERMYRP 146
QY 517 NVVKLGKYYLFAATRLNRGSDNDAMNANYAVGD-----NVAMGVYVADSLTGSYK- 568
Db 147 DAIOQL-----AIIITDGIQND--LPRSTAVVHOLKRRKHHVVAIIIG-VGAGVNNENYR 195
QY 569 -----PLNDSGVVLTASVPANWRTATSYAVVPVAGKDDQVLTYSYMTNRNGVA 617
Db 196 ILVGCDDRYAPCPYSSG-----SNSEA-----QNNIKPFELT-----226
QY 618 GKGMDSTWAPSFLLQNPONTTTLVAKMTNQGDW-IMDDSENLD-----661
Db 227 -----KVCQEVERIAHCGKWEWSECSTTCDEGKRIRRRILHPGC 267
QY 662 -----MIGDLSAAL-----PGEROKPVDWDLIGYGLKPHDPATNDPETP 702
Db 268 VSEMTTPCKVRDCQIPIPPVPIPNKIPEKPSNPEEPVN-----PNDPNDPNPNP 318
QY 703 TTPETPTNTPKTPKTPENGTQTPNTPTPEIPTPKPTPKPTQPTNNRLPOTGNNA 762
Db 319 NPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 378
QY 763 N 763
Db 379 N 379
```

Search completed: September 21, 2002, 23:37:39  
Job time: 5080 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 21, 2002, 23:32:00 ; Search time 7643.51 Seconds  
(without alignments)  
4179.663 Million cell updates/sec

Title: US-09-995-587A-2

Perfect score: 2367

Sequence: 1 atgtataaaacgcgtataaa

Scoring table: IDENTITY\_NUC .....ttaacaaacgtcatttaac 2367

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST.\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pin:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	81.2	3.4	989	12 CNS02HA4	AL197365 Tetraodon
2	76.2	3.2	501	12 FR004173	AL444958 Fugu rubr
3	75.4	3.2	735	12 CNS04NSM	AL299119 Tetraodon
4	74.8	3.2	773	12 CNS01VTG	AL169549 Tetraodon
5	72.6	3.1	494	12 CNS048073	AL444858 Fugu rubr
6	71.8	3.0	450	12 FR0025683	AL018519 F. rubripe
7	68.2	2.9	470	12 FR0018463	290754 F. rubripes
8	67.2	2.8	619	12 CNS006944	AL281759 Tetraodon
9	67.2	2.8	806	12 CNS04AEE	AL062806 Drosophila
10	65.6	2.8	959	12 CNS00655	AL444288 Fugu rubr
11	64	2.7	619	12 FR0047503	BM358689 GA_Ea001
12	63.6	2.7	698	10 BM358689	AL444366 Fugu rubr
13	63.6	2.7	979	12 CNS06PDW	AL409242 T7 end of
14	63.4	2.7	619	12 CNS06PDW	AL444386 Fugu rubr
15	63.2	2.7	641	12 FR0047601	AQ946120 Sheared D
16	62.8	2.7	1007	12 CNS06X9S	AL419462 T3 end of
17	62.6	2.6	515	10 C90898	C90898 C90898 Dict

18	62.6	2.6	925	10 BG441163	BG441163 GA_Ea001
19	62	2.6	619	12 FR0013713	AL004959 F. rubripe
20	61.8	2.6	550	12 FR0043207	AL130699 Fugu rubr
21	61.6	2.6	824	12 AZ185454	AZ185454 SP_1005_A
22	60.8	2.6	519	10 BG226324	BG226324 KQ20C02.y
23	60.8	2.6	919	12 BH138523	BH138523 ENT0807TIR
24	60	2.5	479	12 BH211305	BH211305 Sml-50G22
25	60	2.5	933	12 AZ204694	AZ204694 SP_0100_A
26	59.8	2.5	534	12 BH203231	BH203231 Sml-57E13
27	59.6	2.5	394	9 AU179547	AU179547 AU179547
28	59.4	2.5	257	12 FR0026883	AL019712 F. rubripe
29	59.4	2.5	273	12 FR0047466	AL444251 Fugu rubr
30	59.4	2.5	836	12 AQ411961	AQ411961 CPG0931A
31	59.4	2.5	1101	12 CNS0000CT	AL05406 Drosophila
32	59.2	2.5	607	9 BE124614	BE124614 EST393649
33	59.2	2.5	687	10 BG447645	BG447645 NF003A12S
34	59.2	2.5	893	12 AZ542140	AZ542140 ENTEN367F
35	58.4	2.5	865	12 AZ685375	AZ685375 ENTKV53TF
36	58.2	2.5	453	9 AU179269	AU179269 AU179269
37	58.2	2.5	610	9 AU180091	AU180091 AU180091
38	57.8	2.4	567	12 FR0006951	AU180168 AU180168
39	57.8	2.4	593	12 CNS00880	Z90761 F. rubripes
40	57.6	2.4	326	9 AI443192	AI443192 Drosophila
41	57.2	2.4	454	9 AI496419	AI496419 sb04e01.y
42	57.2	2.4	474	10 BF595475	BF595475 SU77F03.y
43	57.2	2.4	488	9 BE020655	BE020655 sm51106.y
44	57.2	2.4	748	12 AZ183618	AZ183618 SP_1001_A
45	56.8	2.4			

## ALIGNMENTS

## RESULT 1

## CNS02HA4

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

CNS02HA4  
Tetraodon nigroviridis genome survey sequence 17 end of clone  
139021 of library G from Tetraodon nigroviridis, genomic survey  
sequence.  
AL197365.1 GI:7835515  
GSS: genome survey sequence.  
Tetraodon nigroviridis.  
Tetraodon nigroviridis.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontidae; Tetraodon.  
1 (bases 1 to 989)  
Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,  
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and  
Weissenbach,J.  
Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetraodon nigroviridis  
Unpublished  
2 (bases 1 to 989)  
Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,  
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,  
Saurin,W. and Weissenbach,J.  
Human gene number estimate provided by genome wide analysis using  
Tetraodon nigroviridis DNA sequence  
Unpublished  
3 (bases 1 to 989)  
Genoscope.  
Direct Submission  
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
<http://www.genoscope.cns.fr/Tetraodon>.  
Location/Qualifiers  
1. .989  
/organism="Tetraodon nigroviridis"





1. (bases 1 to 773)  
Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Wetzenbach, J.  
Characterization and repeat analysis of the compact genome of the freshwater pufferfish *Tetraodon nigroviridis*  
unpublished
2. (bases 1 to 773)

Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence  
unpublished  
3 (bases 1 to 773)

This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

```

4. 173
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="199C24"
/clone_lib="C"

```

### h Similarity

ACTCATGATACTGCTACTACTACWACTGCTACWACTAGTCWACWAGCTACTGCWA 213

*ggttaaaaagtgtcacacttggtagtggtcaagtttagtcagctagtcatacga*tatta 211  
||||| :||: ||: | :||: ||| :||| ||: |  
*GGTACTGCTACTACWAGTGCWAGTATTCCTCAATCCCATCCTCGTTGGT*

ACTACTACTACTRCTACTACTACTGCAWACCAGCATACATACTACTACTACT 393  
tagcaagttctgctgcaataacatcatctacaagttccgcagotcattcaataaca 331

CTACWRCWRCACACTACTGCTACAACACTACTACTRCWACTGCTACTACTACTACTA 513

cagctgcaaaaaatgatgatcaacaatccagtaaaaagaatactaccgcta 505  
| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
TTTCTGTAAGCAATCGAATGGTAAACGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT

R0048073 494 bp DNA linear GSS 05-JAN-2001

```

LOCUS      FR0025683
DEFINITION F.rubripes GSS sequence, clone 154E17ac12, genomic survey sequence.
ACCESSION  AL018519
VERSION    AL018519.1 GI:2684887
KEYWORDS   GSS; genome survey sequence.
SOURCE     Takifugu rubripes.
ORGANISM   Takifugu rubripes.
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Takifugu.
REFERENCE  1 (bases 1 to 450)
            Elgat,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrانيا,Y.,
            Williams,G. and Brenner,S.
TITLE      Direct Submission
JOURNAL    Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource
            Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hmp.mrc.ac.uk
COMMENT    V type: phagemid
            PRIMER: KS
DESCR:     One pass dye-terminator sequencing of cosmid cloned genomic
            sequence.

FEATURES             Location/Qualifiers
     1..450
         /organism="Takifugu rubripes"
         /db_xref="taxon:31033"
         /clone_lib="cosmid 154E17"
         /clone="154E17ac12"
     96 a 140 c 64 g 144 t 6 others
BASE COUNT
ORIGIN

```

Query Match	3.0%;	Score 71.8;	DB 12;	Length 450;	
Best Local Similarity	50.18;	Pred. No. 1.6e-05;			
Matches 178;	Conservative	0;	Mismatches 177;	Indels	0; Gaps 0;

  

QY	66	aacaaactgtaaatgcgtccgagcacacaataattgaaacaaatgattctctactgtaca	125
Db	61	ACTCTCTAATGTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC	120
QY	126	agttacaaacaggtgataatgatattgctgtttaaaagttgacacttggtagtggccaagt	185
Db	121	TAC	180
QY	186	tagtcagctagtcatacgaactattagaactcttgcataatgcacaaatagctctctctgc	245
Db	181	TACTGCTACTACTGCTACTGCTACTGCTACTGCTACTATTACTACTGCTACTACTGCTGTC	240
QY	246	cqctaatacacaaaattctaacagtcgaagtagcaagttctgctgcataacaatcatctac	305
Db	241	TGCTACTACTGCTACTACTACTGCTACTACTGCTACTGCTGCTGCTGCTGCTGCTGCTAC	300
QY	306	aagttccgcagcttctaataataacacagatagtaaaagcggtcgaagaaatacataatc	365
Db	301	TGCTGCTGCTACTGCTACTATTACTACTGCTACTACTGCTGCTGCTACTACTGCTACTAC	360
QY	366	agccaaaaatgatcacacgcgcaaaaagctgcaccagctcaagaaactctctgaagct	420
Db	361	TACTGCTACTGCTGCTGCTACTACTGCTACTACTGCTACTGCTACTGCTGCTGCTGCT	415

  

RESULT 7	
FR0018463/c	
LOCUS	470 bp DNA linear
DEFINITION	F.rubripes GSS sequence, clone 016R10Ac6, genomic survey sequence.
ACCESSION	AL011359
VERSION	AL011359.1 GI:2676703

  

FR0018463	470 bp	DNA	linear	GSS 09-DEC-1997
F.rubripes	GSS sequence, clone 016R10Ac6,	genomic survey sequence.		
AL011359				
AL011359.1	GI:2676703			

SOURCE	ORGANISM
Takifugu rubripes.	
Takifugu rubripes.	
Takifugu rubripes.	
Eukaryota; Metazoa:	
Actinopterygii; Neopterygii;	Craniata: Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii;	Teleostei; Euteleostei; Neoteleostei;

REFERENCE	Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Takifugu.
AUTHORS	1 (bases 1 to 470) Elgar,G., Clark,M., Smth,S., Meek,S., Warner,S., Umrania,Y., Williams,G. and Brenner,S.
TITLE	Direct Submission
COMMENT	Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgm.mrc.ac.uk V_type: phagemid PRIMER: KS DESCR: One pass dye-terminator sequencing of cosmid cloned genomic sequence.

RESULT 8  
FR0006944  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
GSS; genome survey sequence.  
Takifugu rubripes.  
Takifugu rubripes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontidae; Takifugu.  
1 (bases 1 to 619)  
Elgar, G., Clark, M., Smith, S., Meek, S., Warner, S., Umrانيا, Y.,  
Williams, G. and Brenner, S.  
Direct Submission  
Submitted (18-FEB-1997) MRC Human Genome Mapping Project Resource  
Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgmrc.ac.uk  
Vector: pBluescript II KS

```

V_type: phagemid
PRIMER: KS
DESCR:
One pass dye-terminator sequencing of cosmid cloned genomic
sequence.
FEATURES             Location/Qualifiers
     source            1..619
                        /organism="Takifugu rubripes"
                        /db_xref="taxon:31033"
                        /clone_lib="cosmid 133Bl6"
                        /clone="133Bl6aC7"
BASE COUNT          75 a   150 c   116 g   175 t   103 others
ORIGIN
0:
Query Match           2.8%; Score 67.2; DB 12; Length 619;
Best Local Similarity 48.3%; Pred. No. 0.00018;
Matches 174; Conservative 0; Mismatches 186; Indels 0; Gaps 0;
OY 63  tgcaacaactgtaaatgcacccgcggcacacaaatttgaacaagaatgattctttactgt 122
Db    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
14  TGCTACTANNACTGCTGTCTACNGCTACTACNGCTNCTGCTACTAATGCTACTGTGCTTAC 73
OY 123 acaaagttaacacagtgataatgatattgctgtttaaaaagtgcaccttgtagtggtca 182
Db    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
74  TGCNACTACTGCTACTGCTGTGCTACTGCTACTGCTACTGCTACTACTACTACTGCTGTC 133
OY 183 agttagtcagctagtgatacgaactatttagaacctctgctaataagcgaatagtccttc 242
Db    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
134  TGCTGTCTGCTGCTGTGCTACTACTGCTGTGCTACTGCTGCTGCTACTGCTGCTGCTACTGC 193
OY 243 tgcgcgctaatacacaattctcaacagtcgaagtccaagttctgtgctgccaatacatcatc 302
Db    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
194  TACTACTACTGCTGCTACTGCTACTGCTACTGCTGTGCTGCTACTACTGCTACTGCTACTAC 253
OY 303 tacaagttccgcagcttcattaaataaacacagatagtaaagcgcgctcaagaaaaactaa 362
Db    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
254  TACTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTACTGCTACTGCTACTGCTACTGA 373
OY 363 tacagccaaaaatgatgacacgcgaaaaagctgaccacgactaacgaatatctctgaagctaa 422
Db    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
314  TACTGCTACTGCTGCTACTANTGNWTAAGTGCTGCTGCTGCTGCTGCTGCTGCTACTGCTACT 473

```

RESULT	9	
LOCUS	CNS044AE/c	
DEFINITION	CNS044AE 806 bp DNA linear GSS 21-MAY-2000 Tetraodon nigroviridis genome survey sequence T7 end of clone 095624 of library G from Tetraodon nigroviridis, genomic survey sequence.	
ACCESSION	AL281759	
VERSION	GI:8020086	
KEYWORDS	GSS: genome survey sequence.	
SOURCE	Tetraodon nigroviridis.	
ORGANISM	Tetraodon nigroviridis	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Acanthomorpha; Acanthopterygii; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.	
REFERENCE	1 (bases 1 to 806)	
AUTHORS	Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.	
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis	
JOURNAL	2 (bases 1 to 806)	
REFERENCE	Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.	
AUTHORS	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence	
TITLE	Unpublished	
JOURNAL	Unpublished	

```

REFERENCE 3 (bases 1 to 806)
AUTHORS   Genoscope.
TITLE     Direct Submission
JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT   This sequence is a single read and was generated as part of a large
          scale clone-end sequencing project of the Tetraodon nigroviridis
          genome. For more information, please take a look at
          http://www.genoscope.cns.fr/tetraodon.

FEATURES             Location/Qualifiers
     source            1..806
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="095G24"
                     /clone.lib="G"
                     /note="Genoscope sequence ID : CORG095BD12LPI-end : T7"

BASE COUNT  247 a 87 c 99 g 333 t 40 others
ORIGIN

Query Match      2.8%; Score 67.2; DB 12; Length 806;
Best Local Similarity 41.38; Pred. No. 0.0002;
Matches 224; Conservative 29; Mismatches 289; Indels 0; Gaps 0;

QY  874  acccaaatgataatcatctcttcttataatgaatgtagtgataatgaatgaattgaatt 933
      || || || || || || || || || || || || || || || || || || || || ||
Db  737  AATAATAATAGTAATAATCAATAAATAATAATAATAATAGTACTACTACTACTACTAATAAT 678

QY  934  catggagaatgtaggtcccaatttttggctataattctaccgcggtttccacaagaatgg 993
      || || || || || || || || || || || || || || || || || || || || ||
Db  677  AATAGTAATAATAGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 618

QY  994  tcaggatcagctgtttgaacagtgataactctcaatctcaattttttatacaaggtagac 1053
      : : : : : || || || || || || || || || || || || || || || || || ||
Db  617  AATAAATAATAATGATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 558

QY  1054  acgtctgataacaataccaatcatcaaaaattgtcagctactctctttttaaactgat 1113
      || || || || || || || || || || || || || || || || || || || || ||
Db  557  AAMSSTATTATATACMAAHMACMAHMMCMSCCMCMCMCMCMCMCMCMCMCMCMCMCMCMCM 498

QY  1114  aataatggaatgtatcactcgcgtcaggtacgaatgactattatttgaaggtgat 1173
      || || || || || || || || || || || || || || || || || || || || ||
Db  497  AATAATAATAATGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 438

QY  1174  gctattactacaacattatgataatggaagctactaacaagaagtgccgataatt 1233
      || || || || || || || || || || || || || || || || || || || || ||
Db  437  AATAATAGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 378

QY  1234  gcaatgcgtgctcatgtaattgaagatggaatggtgtagctggtacacctgttttgaa 1293
      || || || || || || || || || || || || || || || || || || || || ||
Db  377  GGTATTAAATGATAGTAATSAAYATAATAGYAYATAATAATAATAATAATAATAATAATA 318

QY  1294  9caagtactggttggaataattatcaaggcgaggaccataatttataactggttaattat 1353
      || || || || || || || || || || || || || || || || || || || || ||
Db  317  AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 258

QY  1354  9ccgagatgacgcatttaataaagacttattagaattctttccaaatgatgatatt 1413
      || || || || || || || || || || || || || || || || || || || || ||
Db  257  AGTAATGATAATCATAGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 198

QY  1414  aa 1415
      ||
Db  197  AA 196

RESULT 10
CNS00655
LOCUS
DEFINITION  CNS00655 959 bp DNA linear GSS 03-JUN-1999
            Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR13F18 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION  AL062806
VERSION   AL062806.1 GI:4943996
KEYWORDS  GSS.

```

```

fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 959)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES             Location/Qualifiers
     source            1..959
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone.lib="RPCI-98"
                     /clone="BACR13F18"
                     /note="end : TET3"

BASE COUNT  525 a 63 c 52 g 106 t 213 others
ORIGIN

Query Match      2.8%; Score 65.6; DB 12; Length 959;
Best Local Similarity 31.9%; Pred. No. 0.00046;
Matches 210; Conservative 111; Mismatches 337; Indels 0; Gaps 0;

QY  291  aatacatcatctacaagttccgcagcttcattataacacagatagtaagcgctca 350
      || || || || || || || || || || || || || || || || || || || || ||
Db  298  AAAAAAATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 357

QY  351  agaaatctactacagcaaaatgatgacacgcaaaagctgcacagctcaagatc 410
      || || || || || || || || || || || || || || || || || || || || ||
Db  358  ANAWTTAAATACACGCGSSSYTWMMMMMMCMMAAASRTTSTTSCCWTCCCCCCCM 417

QY  411  ttctgaagctaaaatgaaccagctgtaaacgttaattgattcttcagctgcaaaaatga 470
      || || || || || || || || || || || || || || || || || || || || ||
Db  418  CCCCSARRAAATAAAAAAATTTTGTGGMTTKTWATGKNAATWAGARACAAAGRAKA 477

QY  471  tgatcaaatccctgaataaagaatactaccgctgaagttaacaggtgctgaaacgt 530
      || || || || || || || || || || || || || || || || || || || || ||
Db  478  RARGAAAAAATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 537

QY  531  tgtaaaaagcgcggaattgatccctacagtttaactgatgaccagattaagcattaaa 590
      || || || || || || || || || || || || || || || || || || || || ||
Db  538  ARWAKDAARAGGGGAATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 597

QY  591  taagatgaactctcgaagctgcaagctggtgtacacaaatgacttaataatgtttcca 650
      || || || || || || || || || || || || || || || || || || || || ||
Db  598  AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 657

QY  651  aaagattgctgatactgtaatacaacagatggtcggtacacagcttcattcttaagc 710
      || || || || || || || || || || || || || || || || || || || || ||
Db  658  AAAMWTTTATTWTTAKRAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 717

QY  711  aagtgaatcaaaatgctgcgcgtacaaactaaagatgcacaaactaactattga 770
      || || || || || || || || || || || || || || || || || || || || ||
Db  718  RDAARRAAAAAAAWADGRKAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 777

```

[illegible]

RESULT	11
FR0047503	
LOCUS	
DEFINITION	619 bp DNA linear GSS 05-JAN-2001 Fugu rubripes GSS sequence, clone 217A13gfl, genomic survey sequence.
ACCESSION	AL444286
VERSION	AL444288.1 GI:12052124
KEYWORDS	GSS; genome survey sequence.
SOURCE	Takifugu rubripes.
ORGANISM	Takifugu rubripes.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Takifugu.

AUTHORS  
Clark M S  
TITLE  
Direct Submission  
JOURNAL  
Submitted (29-SEP-2000) MRC Human Genome Mapping Project Resource  
Centre Hinxton Cambridge, CB10 1SB, UK Email:  
biohelp@hmp.mrc.ac.uk  
COMMENT  
vector: pBluescript II KS  
V.type: phagemid  
PRIMER: KS

**FEATURES**

One pass dye-terminator sequencing of BAC (pBelOBACII) cloned genomic sequence  
The BACs can be obtained from <http://www.incyte.com>.  
Location: 900115.

```

source
location/Qualifiers
1. .619
/organism="Takifugu rubripes"
/db_xref="taxon:31033"
/clone="217A13gfl1"
/clone-lib="BAC 217A13"
102 a 161 c 130 g 143 t 83 others
BASE COUNT
ORIGIN

```

Query Match	2.7%;	Score 64;	DB 12;	Length 619;
Best Local Similarity	50.4%;	Pred. No. 0.00092;		
Matches 142;	Conservative	0;	Mismatches 146;	Indels 0;
Gaps	0;			
67	acaactgtaaaatgcattccgcggacacaataattgaaaaacaatgattcttctactcgtacaa	126		
b	1	ACTACTACTACTACTACTGCTACTACTACTACTACTACTGCTGCTTCTACTACTGCTACT	60	
127	gttaacaacagggtgataatgatattcgtctgtttaaagcgtgcacattggtagtgggtcaagtt	186		
b	61	ACTACTACTACTACTACTACTACTACTGCTGCTTCTACTACTACTACTACTACTGCTACT	120	
187	agtcgaigttagtgatacagactattagaactctctgctaatagcacaatagtgcttctctgcc	246		
b	121	ACTACTACTGCTTCTACTACTACTACTACTACTGCTGCTGCTTCTACTACTGCTACTACT	180	
247	gctaatacacaaaaattctaacacgctcaagtagcaggtctctcgtcgaataaacaatctaca	306		
b	181	ACTACTACNGCTNCTACTACTGCTNCTTCTACTACTACTCTCTACTACTNACTACTACTACT	240	
307	agttccgcagcttcaataaaataacacagatagtaaaaggct	348		
241	ACTNCTGCTGCTGCTNCTNCTNCTACTACTCTCTNCTACTACTCTCTACTACTCTCTACT	282		

RESULT	12
BM358689	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	

[illegible]

## FEATURES

**BASE COUNT  
ORIGIN**

	Query Match	2.7%;	Score 63.6;	DB 10;	Length 698;
	Best Local Similarity	59.3%;	Pred. No.	0.0012;	
	Matches 108;	Conservative	0;	Mismatches 74;	Indels 0; Gaps 0;
QY	2064	aagaccgcgatgctcgtctacacacaatgatctctgaacgcgcaactaccagaaaacccc	2123		
Db	36	AANACCCAAAGAACCTTGAGAAGCCTTAAAGTACCAGAAAAACCCAAAAGAACCCGAAAAAGCC	95		
QY	2124	tggagacacctatactctcccaaaacaccaaagactccttgaaaaatccttggagcacctccaac	2183		
Db	96	TAAAGCACCCAGAAAAAACC AAAAGAACCCGAAAAACCCGAGAAACTTAAACA	155		
QY	2184	tctaatacacctaatactccggagaattcctttaactccgaaacgcctaagcaaacctga	2243		
Db	156	ACCCGNNAAAGCGGAATGAACCCCGAAAAACCCAAAAGAACCCGAGAAACCCGAAACCCCA	215		
QY	2244	aa	2245		
Db	216	AA	217		

RESULT	13
CNS06PDW/c	
LOCUS	
DEFINITION	CNS06PDW
	T7 end of clone AV0AA013G05 of library AV0AA from strain CBS 379 of
	979 bp DNA linear GSS 05-JUL-2001

QY	295	acatcatctacaagttccgcagcgtcttcaattataataacacagatagataaaagcggctcaagaa	354
Db	168	ACCGATCAACAGATGGTAACGGTTCCTCCCAACACAGATGATGGTAATGGTACACATACG	109
QY	355	aatactaatacagccaaaatgatgatcacgcacaaaagctgca	396
Db	108	ACAACTAATGGTGGTATGGTAAATGAGATGGGATCATCA	67
RESULT	14		
FR0047601			
LOCUS			
DEFINITION			
ACCESSION	FR0047601		
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
BASE COUNT			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
QY	129	tacaacaggtgataatgatattgctgttaaaagtgtacaccttggtggtcaagttag	188
Db	9	TGCAAGCACTACTACTACTGNTACTACTACTACTGCTTCTACTACTACTACTACTACTAC	68
QY	189	tgacgtagtagtaacgactattagaactcttgcataatgcataatagtgctcttcttcgcgc	248
Db	69	TGCTACTACTACTACTGCTTCTACTACTACTACTACTACTACTACTACTACTACTACTGC	128
QY	249	taatacacaaaattctaacagtcgaagtctgcgtgcaataacatcatctacaag	308
Db	129	TAC	188
QY	309	ttccgcagcttcattataacacagatagtaaaagcgtcgaataactaataacagc	368
Db	189	TACTACTACTACTGCTTCTACTACTACTACTACTACTACTACTACTACTACTACTACTAN	248
QY	369	caaaaatgatgacacgcaaaagctgcacagcttaacatctcttggaagctaaaaatga	428
Db	249	TACTACTGCTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	428
QY	429	accagcttaaacgcttaatgatattcttcagctgc	461

Search completed: September 22, 2002, 01:51:26  
Job time: 8366 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

\* Run on: September 21, 2002, 23:37:45 ; Search time 180.95 Seconds  
(without alignments)  
3213.122 Million cell updates/sec

Title: US-09-995-587A-2  
Perfect score: 2367

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	80	3.4	10317	3	US-09-058-746-1
C 2	80	3.4	10317	4	US-09-438-142-1
3	68.8	2.9	3730	1	US-08-242-932-8
4	68.8	2.9	3730	1	US-08-714-481-8
5	68.8	2.9	3730	5	PCT-US95-06111-8
6	67.4	2.8	3305	3	US-09-068-043-1
7	64.2	2.7	3294	4	US-08-923-992A-7
8	64.2	2.7	3312	4	US-08-923-992A-3
9	64.2	2.7	3384	4	US-08-923-992A-5
10	64.2	2.7	3492	4	US-08-923-992A-9
11	64.2	2.7	4200	1	US-08-242-932-1
12	64.2	2.7	4200	1	US-08-714-481-1
13	64.2	2.7	4200	5	PCT-US95-06111-1
14	63.8	2.7	4200	4	US-08-923-992A-1
15	63.2	2.7	4673	1	US-07-638-431-1
16	63.2	2.7	4765	5	PCT-US92-00018-1
17	52	2.2	1365	2	US-08-870-827-4
18	52	2.2	2408	2	US-08-870-827-5
19	52	2.2	7218	1	US-08-232-463-14
20	51.8	2.2	3489	2	US-08-728-323A-1
21	51.8	2.2	32207	2	US-08-770-379-20
22	51.8	2.2	32207	4	US-08-757-669A-20
23	51.8	2.2	32207	4	US-09-230-371A-20
24	50.4	2.1	3666	2	US-08-682-517-13
25	50.4	2.1	3666	2	US-08-682-517-14
26	50.4	2.1	4197	2	US-08-682-517-7
27	50.4	2.1	4197	2	US-08-682-517-8

C 28 50 2.1 5102 1 US-08-494-168-1 Sequence 1, Appli  
29 48.8 2.1 5163 3 US-08-700-651-1 Sequence 1, Appli  
30 48.8 2.1 5163 3 US-08-928-361B-4 Sequence 4, Appli  
31 48.8 2.1 5318 3 US-08-700-651-2 Sequence 2, Appli  
32 48.8 2.1 5318 3 US-08-928-361B-3 Sequence 3, Appli  
33 48.4 2.0 1079 1 US-07-781-355-1 Sequence 1, Appli  
34 48 2.0 921 2 US-08-795-475-2 Sequence 2, Appli  
35 48 2.0 1308 2 US-08-795-475-4 Sequence 4, Appli  
36 47.6 2.0 3763 1 US-07-792-865D-1 Sequence 1, Appli  
37 47 2.0 1924 3 US-08-961-083-159 Sequence 159, App  
38 46.4 2.0 19124 2 US-08-487-826B-13 Sequence 13, Appl  
39 46.2 2.0 1725 1 US-08-257-073-14 Sequence 14, Appl  
40 46 1.9 1956 4 US-08-559-896B-1 Sequence 1, Appli  
41 46 1.9 5181 1 US-08-257-073-10 Sequence 10, Appli  
42 45.6 1.9 5511 3 US-08-928-361B-2 Sequence 2, Appli  
43 45.6 1.9 7334 3 US-08-928-361B-1 Sequence 1, Appli  
44 45.6 1.9 8920 2 US-08-446-855A-1 Sequence 1, Appli  
45 45.6 1.9 8920 4 US-09-150-741-1 Sequence 1, Appli

## ALIGNMENTS

RESULT 1  
US-09-058-746-1/c  
; Sequence 1, Application US/09058746  
; Patent No. 6022716  
; GENERAL INFORMATION:  
; APPLICANT: Ilya Chumakov  
; APPLICANT: Hiroaki Tanaka  
; TITLE OF INVENTION: High Throughput DNA sequencing vector  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: Knobbe, Martens, Olson & Bear  
; STREET: 501 West Broadway  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92101-3505  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy Disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: Win95  
; SOFTWARE: Word  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/058,746  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Israel, Ned A.  
; REGISTRATION NUMBER: 29,655  
; REFERENCE/DOCKET NUMBER: GENSET.015A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 235-8550  
; TELEFAX: (619) 235-0176  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10317 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: DOUBLE  
; TOPOLOGY: CIRCULAR  
; MOLECULE TYPE: synthetic DNA  
; ORIGINAL SOURCE:  
; ORGANISM: Cloning vector pGendel  
; FEATURE:  
; NAME/KEY: pGendel  
; LOCATION: 1..10317  
; FEATURE:  
; NAME/KEY: Homology with X06404 compl (411..1668)  
; LOCATION: 9..1266  
; IDENTIFICATION METHOD: blastn against X06404  
; FEATURE:  
; NAME/KEY: Kanamycin resistance gene CDS

```

LOCATION: 142..957
IDENTIFICATION METHOD: By homology to X06404
FEATURE:
NAME/KEY: Tn1000's right end
LOCATION: complement 1332..1371
IDENTIFICATION METHOD: blastn against x60200)
FEATURE:
NAME/KEY: Homology with U46017 (1-472)
LOCATION: 1423..1894
IDENTIFICATION METHOD: blastn against U46017
FEATURE:
NAME/KEY: single stranded DNA replication origin
LOCATION: 1423..1894
IDENTIFICATION METHOD: By homology to U46017
OTHER INFORMATION: mutation T -> C 1658
FEATURE:
NAME/KEY: Homology with U51113 (2382..6997)
LOCATION: 1896..6544
IDENTIFICATION METHOD: blastn against U51113
FEATURE:
NAME/KEY: OriS
LOCATION: 1972..2188
IDENTIFICATION METHOD: By homology to U51113
FEATURE:
NAME/KEY: repELR
LOCATION: 2897..2918
OTHER INFORMATION: Described in seqID 16
FEATURE:
NAME/KEY: RepE
LOCATION: 2903..3034
IDENTIFICATION METHOD: By homology to U51113
FEATURE:
NAME/KEY: T3
LOCATION: 3043..3059
OTHER INFORMATION: Described in seqID 17
FEATURE:
NAME/KEY: LRT3RA
LOCATION: complement 3045..3069
OTHER INFORMATION: Described in seqID 15
FEATURE:
NAME/KEY: IncC
LOCATION: 3070..3320
IDENTIFICATION METHOD: By homology to U51113
OTHER INFORMATION: Insertion 33 bases 3038..3071
FEATURE:
NAME/KEY: Para
LOCATION: 3655..4821
IDENTIFICATION METHOD: By homology to U51113
OTHER INFORMATION: mutation G -> A 3878
FEATURE:
NAME/KEY: ParB
LOCATION: 4821..5792
IDENTIFICATION METHOD: By homology to U51113
FEATURE:
NAME/KEY: ParC
LOCATION: 5865..6382
IDENTIFICATION METHOD: By homology to U51113
FEATURE:
NAME/KEY: Homology with J01688 (complement 175..819)
LOCATION: 6574..7218
IDENTIFICATION METHOD: blastn against J01688
OTHER INFORMATION: mutation A -> G 7096
FEATURE:
NAME/KEY: CDS streptomycin sensitivity gene
LOCATION: complement 6716..7090
IDENTIFICATION METHOD: By homology to J01688
OTHER INFORMATION: mutation A -> G 6728
OTHER INFORMATION: mutation G -> C 6821
OTHER INFORMATION: mutation T -> C 7013
OTHER INFORMATION: mutation T -> A 7058
FEATURE:
NAME/KEY: rpsLR

```

```

LOCATION: 7155..7174
OTHER INFORMATION: Described in seqID 12
FEATURE:
NAME/KEY: SP6
LOCATION: 7230..7248
OTHER INFORMATION: Described in seqID 13
FEATURE:
NAME/KEY: Tn1000's left end
LOCATION: 7252..7291
IDENTIFICATION METHOD: blast (x60200)
FEATURE:
NAME/KEY: Homology with X02730 (complement 37..1959)
LOCATION: 7305..9227
IDENTIFICATION METHOD: blastn against X02730
FEATURE:
NAME/KEY: CDS levansucrase gene
LOCATION: complement 7379..8800
IDENTIFICATION METHOD: By homology to X02730
OTHER INFORMATION: mutation T -> C 7466
OTHER INFORMATION: mutation A -> G 7739
OTHER INFORMATION: mutation T -> C (Asn -> Asp) 8347
OTHER INFORMATION: mutation T -> C 8600
OTHER INFORMATION: mutation G -> A (Ala -> Val) 8772
FEATURE:
NAME/KEY: SLR3
LOCATION: 8711..8731
OTHER INFORMATION: Described in seqID 14
FEATURE:
NAME/KEY: Homology with J01636 (complement 1158..1465)
LOCATION: 9298..9623
IDENTIFICATION METHOD: blastn against J01636
FEATURE:
NAME/KEY: CDS alpha peptide beta-galactosidase
LOCATION: complement 9276..9497
IDENTIFICATION METHOD: By homology to J01636
FEATURE:
NAME/KEY: primer HE1
LOCATION: complement 9465..9479
FEATURE:
NAME/KEY: primer HE2
LOCATION: 9461..9475
FEATURE:
NAME/KEY: primer LacIRS2Avr
LOCATION: complement 9603..9630
FEATURE:
NAME/KEY: primer LacE2Mlu
LOCATION: 9289..9314
FEATURE:
NAME/KEY: Homology with M77789 (1889..2576)
LOCATION: 9629..10315
IDENTIFICATION METHOD: blastn against M77789
FEATURE:
NAME/KEY: high copy-number double-stranded DNA replication origin
LOCATION: complement 9629..10315
IDENTIFICATION METHOD: By homology to M77789
OTHER INFORMATION: mutation C -> T 9803
OTHER INFORMATION: site ScaI 10029 - 10034
OTHER INFORMATION: site PmlI 10038 - 10043
OTHER INFORMATION: CLONING SITES 10031 - 10041
FEATURE:
NAME/KEY: oriLRd
LOCATION: 9856..9881
OTHER INFORMATION: Described in seqID 8
FEATURE:
NAME/KEY: OSI
LOCATION: 10009..10026
OTHER INFORMATION: Described in seqID 10
FEATURE:
NAME/KEY: ORI
LOCATION: complement 10046..10062
OTHER INFORMATION: Described in seqID 11
FEATURE:
NAME/KEY: oriLR

```

LOCATION: complement 10182..10202  
OTHER INFORMATION: Described in seqid 9  
US-09-058-746-1

## Query Match

Best Local Similarity 3.4%; Score 80; DB 3; Length 10317;  
Matches 464; Conservative 0; Mismatches 460; Indels 48; Gaps 6;

```
Qy 983 cacaagaatgctcaggaatcagctgttttgaacagtgataactctatccaattatttata 1042
Db 8322 CACAAGAATGCTCAGGTTGACGCCACATTTACATCTCGGAAAAATCCGTTTATTCTACA 8263
Qy 1043 caagggtagacagctctgataaacaataccaatcatcaaaaaattgtagcgtactcttt 1102
Db 8262 C-----TGATTCTCGGTAACATTTACGGCAACAAACACACGACAACTGCAACAAGTTA 8209
Qy 1103 atttaactgataataatggaatgtatcactcgtcctcaggtacgaaatgactatattgtat 1162
Db 8208 ACGTATCAGCATCAGACAGCTCTTTGAACATCAACGCTGTAGAGGATTTAAATCAATCT 8149
Qy 1163 ttgaagtgatggtctattactacaaacttatgatacaatgga---aagctactacaagaag 1219
Db 8148 TTGAGGCTGACGGAACAACTATCAAAATGTACAGCAGTTCTATCGATGAAGGCAACTACA 8089
Qy 1220 gtccgataatatgcaatgctgtatgctcatgtaattgaagatggttaatggtatcggt 1279
Db 8088 GCTCAGCGCAACACCATACGCTGAGAGATCCTCACTACGTAGATGAAGATGAAGGCCCAAT 8029
Qy 1280 acctgttttgaagaactgactggtttggaataattcaagcgcgaggaacaaatttata 1339
Db 8028 ACTTAGTATTGAGCAACAACTGGAACATGAAGATGGCTTACCAAGCGCAAGATCTTTAT 7969
Qy 1340 actggttaattatgctgagatgacgcatatttaatacaagagcttatttgaattcttt 1399
Db 7968 TTAACAAAGCATACTATGCAAAACGACATCATCTTCGTCAGAAAGTCAAAAACCTC 7909
Qy 1400 ccaatgatgataatgaagtcgagcgaacttgggttaatacgagctatcggtatctctcaaac 1459
Db 7908 TCCAAGCGATAAAACGACGCGGTGAGTTAGCAACGCGCTCTCGGTATGATTGAGC 7849
Qy 1460 taataagagcgaagaatactcctaaagtggtgagagttatatactaccatttaattctgcac 1519
Db 7848 TAAACGATGATTACACTGAAAAGTGATGAA-----ACCGTGATTGCATCTA 7798
Qy 1520 caatgtaagcgaataatgagcgaacaaatgtagttaattaggttaataaattact 1579
Db 7797 ACACGTAACAGATGAATAATGAACGCGGACGCTCTTAAATGAACGCAATGGTATC 7738
Qy 1580 tatttgcgcgtaccgcttttaaatcgaggaagtaagatgatgattggtgaatgtaatt 1639
Db 7737 TGTTCACTGACTCCCGCGGATCAAAATGCGATTGACGGCATTACGCTCAACGATATT 7678
Qy 1640 atgcccgttggtgataatgctgcaatggtcgcgataatgctgctgtagcttaactggtatctt 1699
Db 7677 A-----CATGCTTGTTATGTTCTTAATTTCTTAATCTTAACTGGCCCAT 7639
Qy 1700 ataaagcattaaatgattctgagtagtcttgactccttctgctgcgaactgacacgctg 1759
Db 7638 ARAAGCGCTGACAAACATGGCCCTTGTTGTTAAATATGGATCTTGATCTCAACGATGTA 7579
Qy 1760 cagcaactttatcatattgctgtcccgcttgccggaagaaatgacaaagatttagtta 1819
Db 7578 CCTTTACTTACTCACACTTCGCTGTACC---TCAACGGAAGGAACAATGCGTGATTA 7522
Qy 1820 ctctatattgactaataagaatggaatgagtcggtggaagaatgattcaacttgggcac 1879
Db 7521 CAGCTATTATGAAACAGAGGATTTCTACGCAGACAAA-----CAATCAACGTTTGGC 7468
Qy 1880 cgaattcttactacaaataaaccccggaataacacactactgttttagctaaataagcta 1939
Db 7467 CGAGCTTCCTGCTGACATCAAGGCAAGCAAAACATCTGTTGTCAAGACAGCATCTCTG 7408
```

```
Qy 1940 atcaaggggatt 1951
Db 7407 AACAAAGGACAAAT 7396

RESULT 2
US-09-438-142-1/c
; Sequence 1, Application US/09438142
; Patent No. 6258571
; GENERAL INFORMATION:
; APPLICANT: Ilya Chumakov
; TITLE OF INVENTION: Hiroaki Tanaka
; NUMBER OF SEQUENCES: High Throughput DNA sequencing vector
; CORRESPONDENCE ADDRESS: 22
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/438,142
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelson, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: GENSET.015C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10317 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: synthetic DNA
; ORIGINAL SOURCE:
; ORGANISM: Cloning vector pGendel
; FEATURE:
; NAME/KEY: pGendel
; LOCATION: 1..10317
; FEATURE:
; NAME/KEY: Homology with X06404 compl (411..1668)
; LOCATION: 9..1266
; IDENTIFICATION METHOD: blastn against X06404
; FEATURE:
; NAME/KEY: Kanamycin resistance gene CDS
; LOCATION: 142..957
; IDENTIFICATION METHOD: By homology to X06404
; FEATURE:
; NAME/KEY: Tn1000's right end
; LOCATION: complement 1332..1371
; IDENTIFICATION METHOD: blastn against X06200)
; FEATURE:
; NAME/KEY: Homology with U46017 (1-472)
; LOCATION: 1423..1894
; IDENTIFICATION METHOD: blastn against U46017
; NAME/KEY: single stranded DNA replication origin
; LOCATION: 1423..1894
; IDENTIFICATION METHOD: By homology to U46017
; OTHER INFORMATION: mutation T -> C 1658
; FEATURE:
; NAME/KEY: Homology with U51113 (2382..6997)
; LOCATION: 1896..6544
; IDENTIFICATION METHOD: blastn against U51113
```

```

FEATURE:
NAME/KEY: Oris
LOCATION: 1972..2188
IDENTIFICATION METHOD: By homology to U51113
FEATURE:
NAME/KEY: repELR
LOCATION: 2897..2918
OTHER INFORMATION: Described in seqID 16
FEATURE:
NAME/KEY: RepE
LOCATION: 2903..3034
IDENTIFICATION METHOD: By homology to U51113
FEATURE:
NAME/KEY: T3
LOCATION: 3043..3059
OTHER INFORMATION: Described in seqID 17
FEATURE:
NAME/KEY: LPT3RA
LOCATION: complement 3045..3069
OTHER INFORMATION: Described in seqID 15
FEATURE:
NAME/KEY: IncC
LOCATION: 3070..3320
IDENTIFICATION METHOD: By homology to U51113
OTHER INFORMATION: Insertion 33 bases 3038..3071
FEATURE:
NAME/KEY: ParA
LOCATION: 3655..4821
IDENTIFICATION METHOD: By homology to U51113
OTHER INFORMATION: mutation G -> A 3878
FEATURE:
NAME/KEY: ParB
LOCATION: 4821..5792
IDENTIFICATION METHOD: By homology to U51113
FEATURE:
NAME/KEY: ParC
LOCATION: 5865..6382
IDENTIFICATION METHOD: By homology to U51113
FEATURE:
NAME/KEY: Homology with J01688 (complement 175..819)
LOCATION: 6574..7218
IDENTIFICATION METHOD: blastn against J01688
OTHER INFORMATION: mutation A -> G 7096
FEATURE:
NAME/KEY: CDS streptomycin sensitivity gene
LOCATION: complement 6716..7090
IDENTIFICATION METHOD: By homology to J01688
OTHER INFORMATION: mutation A -> G 6728
OTHER INFORMATION: mutation G -> C 6821
OTHER INFORMATION: mutation C -> T 6866
OTHER INFORMATION: mutation T -> C 7013
OTHER INFORMATION: mutation T -> A 7058
FEATURE:
NAME/KEY: rpsLR
LOCATION: 7155..7174
OTHER INFORMATION: Described in seqID 12
FEATURE:
NAME/KEY: SP6
LOCATION: 7230..7248
OTHER INFORMATION: Described in seqID 13
FEATURE:
NAME/KEY: Tn1000's left end
LOCATION: 7252..7291
IDENTIFICATION METHOD: blast (x60200)
FEATURE:
NAME/KEY: Homology with X02730 (complement 37..1959)
LOCATION: 7305..9227
IDENTIFICATION METHOD: blastn against X02730
FEATURE:
NAME/KEY: CDS levansucrase gene
LOCATION: complement 7379..8800
IDENTIFICATION METHOD: By homology to X02730
OTHER INFORMATION: mutation T -> C 7466

```

```

OTHER INFORMATION: mutation A -> G 7739
OTHER INFORMATION: mutation T -> C (Asn -> Asp) 8347
OTHER INFORMATION: mutation T -> C 8600
OTHER INFORMATION: mutation G -> A (Ala -> Val) 8772
FEATURE:
NAME/KEY: SLR3
LOCATION: 8711..8731
OTHER INFORMATION: Described in seqID 14
FEATURE:
NAME/KEY: Homology with J01636 (complement 1158..1465)
LOCATION: 9298..9623
IDENTIFICATION METHOD: blastn against J01636
FEATURE:
NAME/KEY: CDS alpha peptide beta-galactosidase
LOCATION: complement 9276..9497
IDENTIFICATION METHOD: By homology to J01636
FEATURE:
NAME/KEY: primer HE1
LOCATION: complement 9465..9479
FEATURE:
NAME/KEY: primer HE2
LOCATION: 9461..9475
FEATURE:
NAME/KEY: primer LacIRS2Avr
LOCATION: complement 9603..9630
FEATURE:
NAME/KEY: primer LacE2M1u
LOCATION: 9289..9314
FEATURE:
NAME/KEY: Homology with M77789 (1889..2576)
LOCATION: 9629..10315
IDENTIFICATION METHOD: blastn against M77789
FEATURE:
NAME/KEY: high copy-number double-stranded DNA replication origin
LOCATION: complement 9629..10315
IDENTIFICATION METHOD: By homology to M77789
OTHER INFORMATION: mutation C -> T 9803
OTHER INFORMATION: site ScaI 10029 - 10034
OTHER INFORMATION: site PmlI 10038 - 10043
OTHER INFORMATION: CLONING SITES 10031 - 10041
FEATURE:
NAME/KEY: oriLrd
LOCATION: 9856..9881
OTHER INFORMATION: Described in seqID 8
FEATURE:
NAME/KEY: OS1
LOCATION: 10009..10026
OTHER INFORMATION: Described in seqID 10
FEATURE:
NAME/KEY: OR1
LOCATION: complement 10046..10062
OTHER INFORMATION: Described in seqID 11
FEATURE:
NAME/KEY: oriLr
LOCATION: complement 10182..10202
OTHER INFORMATION: Described in seqID 9
US-09-438-142-1

```

```

Query Match      3.4%; Score 80; DB 4; Length 10317;
Best Local Similarity 47.7%; Pred. No. 4e-11;
Matches 464; Conservative 0; Mismatches 460; Indels 48; Gaps 6;

QY 983 cacaagaatggtcagatcagctgttttgacagcgtgataactctatccaatttttata 1042
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8322 CACAAGATGTCAGGTTTCAGCCACATTACATCTGACGAGAAAATCGTTTATTCTACA 8263

QY 1043 caagggtagacacgtctgtatacaataaccatcatcaaaaaattgctagcgtactcttt 1102
      || || || || || || || || || || || || || || || || || || || ||
Db 8262 C-----TGATTTCCTCCGGTAAACATTACGGCAACAACACTGACAACAGTTA 8209

QY 1103 atttaactgataataatggaatgtatcactcgtcaggtacgaaatgactatattgat 1162
      | || | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Db 8208 ACGTATCAGCATCAGACAGCTCTTTGACATCAACGGTGTAGAGGATTATAAATCAATCT 8149
Qy 1163 ttgaagggtgagcctattactaccacaaacttatgatcaatgga-----aagclactaacaagaag 1219
Db 8148 TTGACGGTGCAGGAAACAGTATCAAAATGTACAGCAGCTTCATCGATGAAGGCAACTACA 8089
Qy 1220 gtgcgcgaataattgcaatgctgctcatgtaattgaatggaatggtgacggt 1279
Db 8088 GCTCAGGCGCAACCATACGCTGAGAGATCTCTACTACGTAGAGATATAAAGGCCACAAAT 8029
Qy 1280 acctgttttgaagcaagtaactggtttggaaaaattatcaagcgagggaccacaaattata 1339
Db 8028 ACTTAGTATTGTAAGCAAAACACAGCTGGAACCTGAAGATGGCTACCAAGCGGAGAAATCTTAT 7969
Qy 1340 actggtttaattggtgagagatgacacatttaataatcaagagcttattagaattcttt 1399
Db 7968 TTAACAAAGCATACTATGSCAAAGACATCATCTCTCCGTCAAGAAAGTCAAAAATCTC 7909
Qy 1400 ccaatgatgatataagagtcggtggaacttggcgtaactgcagctatcgctatcctcaaac 1459
Db 7908 TGCAAAAGCGATATAAAGAACGACGCTGAGTTAGCAAAACGCGCTCTCGGTATGATGAGC 7849
Qy 1460 taataagagcaaaagaatacttaagtggtgagaggtatatacaccattatttctgcac 1519
Db 7848 TAAACGATGATACACACTGAAAGAAAGTATGAA-----ACCGCTGATTCGATATTT 7798
Qy 1520 caatggtgaagcgatgaattgagcgacaaatgtagttaaattaggttaataataattact 1579
Db 7797 ACACAGTAAACAGATGAAATTTGAACGCGGAAAGCTCTTTAAATGAAACGGCAATGGTATC 7738
Qy 1580 tatttcgcctaccgttttaataatgaggaagtaataatgatgctgtggaatgaatgaatt 1639
Db 7737 TGTTCACGTACTCCCGCGGATCAAAAATGACGATTTGACGGCATTTACGTCTAACGATATTT 7678
Qy 1640 atgctgtgtgataatgttgaatgttcggtatgttgcgataatgttgcgataatgaatgaat 1699
Db 7677 A-----CATGCTGTGTATGTTTCTTAATCTTTTAATCTTTTAATCTTTTAATCTTT 7639
Qy 1700 atagccattaaatgattctgagtagtcttgactgtcttcttcttcttcttcttcttcttct 7579
Db 7638 ACAGCGCTGACAAACACTGGCTTGTGTAAATATGGATCTTTGATCTTAAACGATGATA 7579
Qy 1760 cagcaacttattattatgtctcccgcttgcggaagaaagatgacaaagtattagttta 1819
Db 7578 CTTTACTTACTACACTTCGCTGTACC---TCAAGCGAAAGGAAACAAATGTCGTGATTA 7522
Qy 1820 ctccatataactaataagaaatggagtagcgggtgtaaaaggaatggattcaacttggcgcac 1879
Db 7521 CAAGCTATATGACAAACAGAGGATTCTACGCAGACAAA-----CAATCAACGTTTGGCG 7468
Qy 1880 cgaagttcttactacaaattacccggataacacaaactactgttttagctaaatgacta 1939
Db 7467 CGAGCTTCCGCTGACATCAAGGCAAGAAACATCTGTTGTCAAAAGACGATCCTTGTG 7408
Qy 1940 atcaaggggatt 1951
Db 7407 AACAGGACAAAT 7396

```

## RESULT 3

```

US-08-242-932-8
; Sequence 8, Application US/08242932
; Patent No. 5595740
; GENERAL INFORMATION:
; APPLICANT: Brady, L. Jeannine
; TITLE OF INVENTION: Cloning of No. 5595740-IgA Fc Binding Forms of
; the Group B Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL

```

```

; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/242,932
; FILING DATE: 16-MAY-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF142
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3730 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-242-932-8

```

```

Query Match 2.9%; Score 68.8; DB 1; Length 3730;
Best Local Similarity 60.9%; Pred. No. 2.1e-08;
Matches 112; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
QY 2065 aaaccgcatgactctgtacacaaatgactcctgaaacgcaactacaccagaaacccct 2124
Db 2609 ACACCGAGCGCTCAGACACACCGCATGTTCCGGAATCACCAAGGCCCCAGAGACCG 2668
QY 2125 gagacacctaatctcccaaaacacacaaagactcctcgaaatctctgggacacctcaact 2184
Db 2669 CGTGTTCGGGAATCACCAAGACTCCAGAGACCGCATGTTCCGGAATCACCAAGACT 2728
QY 2185 cctaatacacctaactcctcggaatctcttaactccagaaacgcttaagcaactgaa 2244
Db 2729 CAGAGACGACCAAGATTCGGAACCCCTTAAGACTCCAGACGCTCCCTTAAGCTTCAGAC 2788
QY 2245 accc 2248
Db 2789 GTCC 2792

```

## RESULT 4

```

US-08-714-481-8
; Sequence 8, Application US/08714481
; Patent No. 5766606
; GENERAL INFORMATION:
; APPLICANT: Brady, L. Jeannine
; TITLE OF INVENTION: Cloning of No. 5766606-IgA Fc Binding Forms of
; the Group B Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/714,481
; FILING DATE: 16-SEP-1996

```

```

; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/242,932
; FILING DATE: 16-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UFI142
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3730 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-714-481-8

Query Match          2.9%; Score 68.8; DB 1; Length 3730;
Best Local Similarity 60.9%; Pred. No. 2.1e-08;
Matches 112; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 2065 aaacgcgatgctctgtacacccaatgatctgaaacgcaactacacccagaaacccct 2124
Db 2609 ACACCGCAGGCTCCAGACACCGCATGTTCCGGAAATCACCAGAGGCCCCAGAGACCG 2668
QY 2125 gagacacctaatactcccaaacacccaagaagctctgtaaaatcctgggacacctcaaaact 2184
Db 2669 CGTGTTCGGAAATCACCAGAACTCCAGAACGCCGATGTTCCGGAAATCACCAGAACT 2728
QY 2185 cctaatacaccctaaactccgggaattcttaactccagaaacgcttaagcaacccctgaa 2244
Db 2729 CCAGAGACCAACCAAGATTCCGGACCCCTAAGACTCCAGACGTCCTTAAGCTTCCAGAC 2788
QY 2245 accc 2248
Db 2789 GTCC 2792

```

## RESULT 5

```

PCT-US95-06111-8
; Sequence 8, Application PC/TUS9506111
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT: Street address: 186 Grinter Hall
; APPLICANT: City: Gainesville
; APPLICANT: State/Province: Florida
; APPLICANT: Country: US
; APPLICANT: Postal code/zip: 32611
; APPLICANT: Phone number: 904-392-8929
; APPLICANT: Fax number: 904-392-6600
; APPLICANT: Telex number:
; TITLE OF INVENTION: Cloning of Non-IGA Fc Binding Forms of
; TITLE OF INVENTION: the Group B Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06111
; FILING DATE:

```

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/242,932
; FILING DATE: 16-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UFI142
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3730 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; PCT-US95-06111-8

```

```

Query Match          2.9%; Score 68.8; DB 5; Length 3730;
Best Local Similarity 60.9%; Pred. No. 2.1e-08;
Matches 112; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 2065 aaacgcgatgctctgtacacccaatgatctgaaacgcaactacacccagaaacccct 2124
Db 2609 ACACCGCAGGCTCCAGACACCGCATGTTCCGGAAATCACCAGAGGCCCCAGAGACCG 2668
QY 2125 gagacacctaatactcccaaacacccaagaagctctgtaaaatcctgggacacctcaaaact 2184
Db 2669 CGTGTTCGGAAATCACCAGAACTCCAGAACGCCGATGTTCCGGAAATCACCAGAACT 2728
QY 2185 cctaatacaccctaaactccgggaattcttaactccagaaacgcttaagcaacccctgaa 2244
Db 2729 CCAGAGACCAACCAAGATTCCGGACCCCTAAGACTCCAGACGTCCTTAAGCTTCCAGAC 2788
QY 2245 accc 2248
Db 2789 GTCC 2792

```

## RESULT 6

```

US-09-068-043-1
; Sequence 1, Application US/09068043
; Patent No. 6048694
; GENERAL INFORMATION:
; APPLICANT: MICHAEL GENE BRAMUCCI
; APPLICANT: VASANTHA NAGARAJAN
; TITLE OF INVENTION: POSITIVE SELECTION
; TITLE OF INVENTION: VECTOR FOR BACILLUS SP.
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND
; ADDRESSEE: COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT OFFICE 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/068.043
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/006,201
; FILING DATE: NOVEMBER 3, 1995
; ATTORNEY/AGENT INFORMATION:

```

Query Match 2.7%; Score 64.2; DB 4; Length 3294;  
Best Local Similarity 53.9%; pred. No. 3.1e-07;  
Matches 132; Conservative 0; Mismatches 113; Indels 0; Gaps

QY 2012 taccctgacgacgtgataaaacccctgttgattgggacttaattggtattggattgataaaacccgc 2071  
 Db 2402 TACCTCAAGCCCGACGATACACCGAGCTCCAGACACACCGCATGTTCCGCAATCACAA 2461  
 QY 2072 atgacctgctacacacaaatgactgaaacgcgaactacacacgaacccctgagacac 2131  
 Db 2462 AGGCCCGAAGACACCGCGTTCGCGAATCACCAGACTCCAGAGACACCCCATGTTTC 2521  
 QY 2132 ctaatactcccaaaacacaaagactcctgaaatcctgggacacactcaaaactcctcaata 2191  
 Db 2522 CGGAATCACAAAGGCCCGAGAACCGCGTTCGGAATCACCAGACTCCAGAGAG 2581  
 QY 2192 caactaatactcggaaattcctttaactccgaaacgcctgaacacactgaaacccaaa 2251  
 Db 2582 CACCGCATGTTCCGGAATCACCAGACTCCAGAGACTCCAGAGACTTCCGGAACCCCTA 2641  
 QY 2252 ctaat 2256  
 Db 2642 AGACT 2646

## RESULT 8

US-08-923-992A-3  
 ; Sequence 3, Application US/08923992A  
 ; Patent No. 6280738  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tai, Joseph Y.  
 ; APPLICANT: Blake, Milan S.  
 ; TITLE OF INVENTION: No. 6280738-IgA Fc Binding Forms of the Group B  
 ; TITLE OF INVENTION: Streptococcal Beta Antigens  
 ; NUMBER OF SEQUENCES: 34  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20005  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/923,992A  
 ; FILING DATE: 05-SEP-1997  
 ; CLASSIFICATION: 536  
 ; PRIOR APPLICATION NUMBER:  
 ; FILING DATE: 06-SEP-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Esmond, Robert W.  
 ; REGISTRATION NUMBER: 32,893  
 ; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 371-2600  
 ; TELEFAX: (202) 371-2540  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3312 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1..3312  
 ; US-08-923-992A-3

Query Match 2.7%; Score 64.2; DB 4; Length 3312;  
 Best Local Similarity 53.9%; Pred. No. 3.1e-07;

Matches 132; Conservative 0; Mismatches 113; Indels 0; Gaps 0;  
 QY 2012 taccctgacgacgtgataaaacccctgttgattgggacttaattggtattggattgataaaacccgc 2071  
 Db 2420 TACCTCAAGCCCGACGATACACCGAGCTCCAGACACACCGCATGTTCCGCAATCACAA 2479  
 QY 2072 atgacctgctacacacaaatgactgaaacgcgaactacacacgaacccctgagacac 2131  
 Db 2480 AGGCCCGAAGACACCGCGTTCGCGAATCACCAGACTCCAGAGACACCCCATGTTTC 2539  
 QY 2132 ctaatactcccaaaacacaaagactcctgaaatcctgggacacactcaaaactcctcaata 2191  
 Db 2540 CGGAATCACAAAGGCCCGAGAACCGCGTTCGGAATCACCAGACTCCAGAGAG 2599  
 QY 2192 caactaatactcggaaattcctttaactccgaaacgcctgaacacactgaaacccaaa 2251  
 Db 2600 CACCGCATGTTCCGGAATCACCAGACTCCAGAGACTCCAGAGACTTCCGGAACCCCTA 2659  
 QY 2252 ctaat 2256  
 Db 2660 AGACT 2664

## RESULT 9

US-08-923-992A-5  
 ; Sequence 5, Application US/08923992A  
 ; Patent No. 6280738  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tai, Joseph Y.  
 ; APPLICANT: Blake, Milan S.  
 ; TITLE OF INVENTION: No. 6280738-IgA Fc Binding Forms of the Group B  
 ; TITLE OF INVENTION: Streptococcal Beta Antigens  
 ; NUMBER OF SEQUENCES: 34  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
 ; STREET: 1100 New York Avenue, N.W., Suite 600  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20005  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/923,992A  
 ; FILING DATE: 05-SEP-1997  
 ; CLASSIFICATION: 536  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 60/024,707  
 ; FILING DATE: 06-SEP-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Esmond, Robert W.  
 ; REGISTRATION NUMBER: 32,893  
 ; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 371-2600  
 ; TELEFAX: (202) 371-2540  
 ; INFORMATION FOR SEQ ID NO: 5:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3384 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1..3384  
 ; US-08-923-992A-5

Query Match 2.7%; Score 64.2; DB 4; Length 3384;



	Best Local Similarity	53.9%;	Pred. No. 3.1e-07;	Mismatches	132;	Conservative	0;	Mismatches	113;	Indels	0;	Gaps	0;
QY	2012	tacctggcggaacgtgataaacctgttgattggagacttaattggttatggattaaaccgc	2071										
Db	2405	TACCTCAAGCCCGCAGATACACGCGAGGTCCAGACACACCGCATGTTCGGGAATCACCAA	2464										
QY	2072	atgatcctgtacacccaatgatcctgaaacgcccaactacacccagaacccctgagacac	2131										
Db	2465	AGGCCCGAAGAAGCAGCGGTTCGGGAATCACAAAGACTCCAGAACGCGCATGTTC	2524										
QY	2132	ctaatactcccacaacaaagactcctgaaaatccttgggacactcaaatcctaata	2191										
Db	2525	CGGAATCACCAAGGCCCGAAGACCAGCGGTTCGGGAATCACAAAGACTCCAGAA	2584										
QY	2192	cacctaatcctcggaatctcttaactccagaaacgccttaagcaaacctgaaacccaaa	2251										
Db	2585	CACCGCATGTTCGGGAATCACCAAGACTCCAGAACGACCAAGATTCCGGAACCCCTA	2644										
QY	2252	ctaata 2256											
Db	2645	AGACT 2649											
RESULT 10													
US-08-923-992A-9													
; Sequence 9, Application US/08923992A													
; Patent No. 6280738													
GENERAL INFORMATION:													
APPLICANT: Tai, Joseph Y.													
APPLICANT: Blake, Milan S.													
TITLE OF INVENTION: No. 6280738-IgA Fc Binding Forms of the Group B													
TITLE OF INVENTION: Streptococcal Beta Antigens													
NUMBER OF SEQUENCES: 34													
CORRESPONDENCE ADDRESS:													
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.													
STREET: 1100 New York Avenue, N.W., Suite 600													
CITY: Washington													
STATE: D.C.													
COUNTRY: USA													
ZIP: 20005													
COMPUTER READABLE FORM:													
MEDIUM TYPE: Floppy disk													
COMPUTER: IBM PC compatible													
OPERATING SYSTEM: PC-DOS/MS-DOS													
SOFTWARE: PatentIn Release #1.0, Version #1.30													
CURRENT APPLICATION DATA:													
APPLICATION NUMBER: US/08/923,992A													
FILING DATE: 05-SEP-1997													
CLASSIFICATION: 536													
PRIOR APPLICATION DATA:													
APPLICATION NUMBER: US 60/024,707													
FILING DATE: 06-SEP-1996													
ATTORNEY/AGENT INFORMATION:													
NAME: Esmond, Robert W.													
REGISTRATION NUMBER: 32,893													
REFERENCE/DOCKET NUMBER: 1438.0140001/RWE													
TELECOMMUNICATION INFORMATION:													
TELEPHONE: (202) 371-2600													
TELEFAX: (202) 371-2540													
INFORMATION FOR SEQ ID NO: 9:													
SEQUENCE CHARACTERISTICS:													
LENGTH: 3492 base pairs													
TYPE: nucleic acid													
STRANDEDNESS: single													
TOPOLOGY: linear													
MOLECULE TYPE: CDNA													
FEATURE:													
NAME/KEY: CDS													
LOCATION: 1..3492													
US-08-923-992A-9													

	Query Match	2.7%;	Score 64.2;	DB 4;	Length 3492;
	Best Local Similarity	53.9%;	Pred. No. 3.1e-07;	Mismatches	132;
	Matches	132;	Conservative	0;	Mismatches 113;
	Indels	0;	Gaps	0;	
QY	2012	tacctggcggaacgtgataaacctgttgattggagacttaattggttatggattaaaccgc	2071		
Db	2513	TACCTCAAGCCCGCAGATACACGCGAGGTCCAGACACACCGCATGTTCGGGAATCACCAA	2572		
QY	2072	atgatcctgtacacccaatgatcctgaaacgcccaactacacccagaacccctgagacac	2131		
Db	2573	AGGCCCGAAGAAGCAGCGGTTCGGGAATCACAAAGACTCCAGAACGCGCATGTTC	2632		
QY	2132	ctaatactcccacaacaaagactcctgaaaatccttgggacactcaaatcctaata	2191		
Db	2633	CCCAATCACCAAGGCCCGCAGAACGACCGGTTCGGGAATCACCAAAACACTCCAGAA	2692		
QY	2192	cacctaatcctcggaatctcttaactccagaaacgccttaagcaaacctgaaacccaaa	2251		
Db	2693	CACCGCATGTTCGGGAATCACCAAGACTCCAGAACGACCAAGATTCCGGAACCCCTA	2752		
QY	2252	ctaata 2256			
Db	2753	AGACT 2757			
RESULT 11					
US-08-242-932-1					
; Sequence 1, Application US/08242932					
; Patent No. 5595740					
GENERAL INFORMATION:					
APPLICANT: Brady, L. Jeannine					
TITLE OF INVENTION: Cloning of No. 5595740-IgA Fc Binding Forms of					
TITLE OF INVENTION: the Group B Streptococcal Beta Antigens					
NUMBER OF SEQUENCES: 8					
CORRESPONDENCE ADDRESS:					
ADDRESSEE: Saliwanchik & Saliwanchik					
STREET: 2421 N.W. 41st Street, Suite A-1					
CITY: Gainesville					
STATE: FL					
COUNTRY: USA					
ZIP: 32606					
COMPUTER READABLE FORM:					
MEDIUM TYPE: Floppy disk					
COMPUTER: IBM PC compatible					
OPERATING SYSTEM: PC-DOS/MS-DOS					
SOFTWARE: PatentIn Release #1.0, Version #1.25					
CURRENT APPLICATION DATA:					
APPLICATION NUMBER: US/08/242,932					
FILING DATE: 16-MAY-1994					
CLASSIFICATION: 435					
ATTORNEY/AGENT INFORMATION:					
NAME: Saliwanchik, David R.					
REGISTRATION NUMBER: 31,794					
REFERENCE/DOCKET NUMBER: UF142					
TELECOMMUNICATION INFORMATION:					
TELEPHONE: 904-375-8100					
TELEFAX: 904-372-5800					
INFORMATION FOR SEQ ID NO: 1:					
SEQUENCE CHARACTERISTICS:					
LENGTH: 4200 base pairs					
TYPE: nucleic acid					
STRANDEDNESS: single					
TOPOLOGY: linear					
MOLECULE TYPE: DNA (genomic)					

Db 2832 TACCTCAAGCCCGAGATACACCGCGCTCCAGAGACACCGGATGTTCCGGAATCACCAA 2891  
 QY 2072 atgactctgtacacacaaatgatcttgaacgccaactacacagaaacccctgagacac 2131  
 Db 2892 AGGCCCCAGAGCAGACCGCGTGTTCGGAATCACCAAAGACTCCAGAAGCACCGCATGTTTC 2951  
 QY 2132 ctaataactcccaaaacaccaaagactccctgaaaaatcctctgggacacctaactcctaata 2191  
 Db 2952 CGGAATCACCAAAGGCCCGAGAGACCGCGTGTTCGGAATCACCAAAGACTCCAGAAG 3011  
 QY 2192 cacttaatactccgaaaatctctttaaactccgaaacgccttaagcaaacctgaaacccaaa 2251  
 Db 3012 CACCGCATGTTCCGGAATCACCAAAGACTCCAGAAGACTCCAGAAGACTCCGGAACCCCTA 3071  
 QY 2252 ctaaat 2256  
 Db 3072 AGACT 3076

RESULT 12  
 US-08-714-481-1  
 ; Sequence 1, Application US/08714481  
 ; Patent No. 5766606  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Brady, L. Jeannine  
 ; TITLE OF INVENTION: Cloning of No. 5766606-IgA Fc Binding Forms of  
 ; TITLE OF INVENTION: the Group B Streptococcal Beta Antigens  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Saliwanchik & Saliwanchik  
 ; STREET: 2421 N.W. 41st Street, Suite A-1  
 ; CITY: Gainesville  
 ; STATE: FL  
 ; COUNTRY: USA  
 ; ZIP: 32606  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/714,481  
 ; FILING DATE: 16-SEP-1996  
 ; CLASSIFICATION: 424  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/242,932  
 ; FILING DATE: 16-MAY-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Saliwanchik, David R.  
 ; REGISTRATION NUMBER: 31,794  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 904-375-8100  
 ; TELEFAX: 904-372-5800  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 4200 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; US-08-714-481-1

Query Match 2.7%; Score 64.2; DB 1; Length 4200;  
 Best Local Similarity 53.9%; Pred. No. 3.4e-07;  
 Matches 132; Conservative 0; Mismatches 113; Indels 0; Gaps 0;  
 QY 2012 tacctggcgaaacgataaacctgttgattggacttaattggttatgattgataaacgcgc 2071  
 Db 2832 TACCTCAAGCCCGAGATACACCGCGCTCCAGAGACACCGGATGTTCCGGAATCACCAA 2891  
 QY 2072 atgactctgtacacacaaatgatcttgaacgccaactacacagaaacccctgagacac 2131

Db 2892 AGGCCCCAGAGCAGACCGCGTGTTCGGAATCACCAAAGACTCCAGAAGCACCGCATGTTTC 2951  
 QY 2132 ctaataactcccaaaacaccaaagactccctgaaaaatcctctgggacacctaactcctaata 2191  
 Db 2952 CGGAATCACCAAAGGCCCGAGAGACCGCGTGTTCGGAATCACCAAAGACTCCAGAAG 3011  
 QY 2192 cacttaatactccgaaaatctctttaaactccgaaacgccttaagcaaacctgaaacccaaa 2251  
 Db 3012 CACCGCATGTTCCGGAATCACCAAAGACTCCAGAAGACTCCAGAAGACTCCGGAACCCCTA 3071  
 QY 2252 ctaaat 2256  
 Db 3072 AGACT 3076

RESULT 13  
 PCT-US95-06111-1  
 ; Sequence 1, Application PC/TUS9506111  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Street address: 186 Grinter Hall  
 ; APPLICANT: City: Gainesville  
 ; APPLICANT: State/province: Florida  
 ; APPLICANT: Country: US  
 ; APPLICANT: Postal code/zip: 32611  
 ; APPLICANT: Phone number: 904-392-8929  
 ; APPLICANT: Fax number: 904-392-6600  
 ; APPLICANT: Telex number:  
 ; TITLE OF INVENTION: Cloning of Non-IgA Fc Binding Forms of  
 ; TITLE OF INVENTION: the Group B Streptococcal Beta Antigens  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Saliwanchik & Saliwanchik  
 ; STREET: 2421 N.W. 41st Street, Suite A-1  
 ; CITY: Gainesville  
 ; STATE: FL  
 ; COUNTRY: USA  
 ; ZIP: 32606  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US95/06111  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/242,932  
 ; FILING DATE: 16-MAY-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Saliwanchik, David R.  
 ; REGISTRATION NUMBER: 31,794  
 ; REFERENCE/DOCKET NUMBER: UF142  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 904-375-8100  
 ; TELEFAX: 904-372-5800  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 4200 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; PCT-US95-06111-1

Query Match 2.7%; Score 64.2; DB 5; Length 4200;  
 Best Local Similarity 53.9%; Pred. No. 3.4e-07;  
 Matches 132; Conservative 0; Mismatches 113; Indels 0; Gaps 0;  
 QY 2012 tacctggcgaaacgataaacctgttgattggacttaattggttatgattgataaacgcgc 2071

Db 2832 TACCTCAAGCCCGAGATACACCGCAGCTCCAGACACACCGCATGTTCCGGAATCACCAA 2891  
QY 2072 atgactctgctacacccaatgatctctgaaagccaaactacacagaaacccctgagacac 2131  
Db 2892 AGGCCCGAGAGACCGCGGTGTTCCGGAATCACCAAGACTCCAGAACCGCGATGTTTC 2951  
QY 2132 ctactactcccaaaacacacacacactcctgaaatcctctggaacacacccctcaactcctaata 2191  
Db 2952 CGGAATCACCAAGGCCCGCAGACACCGCGGTGTTCCGGAATCACCAAGACTCCAGAA 3011  
QY 2192 cacttaatactccggaatctccttaactccagaaagcgccttaagcgaacccctgaaacccaaa 2251  
Db 3012 CACCGCATGTTCCGGAATCACCAAGACTCCAGAACCGCGATGTTCCGGAATCACCA 3071  
QY 2252 ctaat 2256  
Db 3072 AGACT 3076

## RESULT 14

US-08-923-992A-1  
; Sequence 1, Application US/08923992A  
; Patent No. 6280738  
; GENERAL INFORMATION:  
; APPLICANT: Tai, Joseph Y.  
; APPLICANT: Blake, Milan S.  
; TITLE OF INVENTION: No. 6280738-IgA Fc Binding Forms of the Group B  
; TITLE OF INVENTION: Streptococcal Beta Antigens  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Avenue, N.W., Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/923.992A  
; FILING DATE: 05-SEP-1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/024,707  
; FILING DATE: 06-SEP-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Esmond, Robert W.  
; REGISTRATION NUMBER: 32,893  
; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4200 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 320..3811  
US-08-923-992A-1

Query Match . 2.7%; Score 63.8; DB 4; Length 4200;  
Best Local Similarity 53.9%; Pred No. 4.3e-07;  
Matches 131; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 2006 ctgctttacctggcgaaacgtgataaacctgttgattgggacttaattggttatggattaa 2065  
Db 2880 CGGAATCACCAAGGCCCGCAGAACCGCGGTGTTCCGGAATCACCAAGACTCCAGAA 2939  
QY 2066 aaccgcatgatcctgctacacccaatgatcctgaaagcgcgaactacacagaaacccctg 2125  
Db 2940 CACCGCATGTTCCGGAATCACCAAGGCCCGCAGAACCGCGGTGTTCCGGAATCACCA 2999  
QY 2126 agacaccttaatactcccaaaacacacacacactcctgaaatcctggaacacacccctcaaac 2185  
Db 3000 AGACTCCAGAACCGCGCATGTTCCGGAATCACCAAGACTCCAGAACCGCGCATGTTCC 3059  
QY 2186 ctaatacacttaatactccggaatctccttaactccagaaacgccttaagcgaacccctgaaa 2245  
Db 3060 CGAACCCCTTAAGACTCCAGACGTCCTTAAGCTTCCACAGCTCCCTTAAGCTTCCAGAG 3119  
QY 2246 ccc 2248  
Db 3120 TCC 3122

## RESULT 15

US-07-638-431-1  
; Sequence 1, Application US/07638431  
; Patent No. 5198535  
; GENERAL INFORMATION:  
; APPLICANT: Hoffman, Stephen L.  
; APPLICANT: Charoenvit, Yupin  
; APPLICANT: Hedstrom, Richard  
; APPLICANT: Khushmith, Srisin  
; APPLICANT: Rogers IV, William O.  
; TITLE OF INVENTION: Protective malaria sporozoite surface protein  
; TITLE OF INVENTION: Immunogen and gene  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: A. David Spevack  
; STREET: NMDC Building 1 T-12 National Naval  
; CITY: Bethesda  
; STATE: MD  
; COUNTRY: USA  
; ZIP: 20814-5044  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/638,431  
; FILING DATE: 19910110  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Spevack, Avrom D.  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 295-6759  
; TELEFAX: (301) 295-4033  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4673 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: N  
; ANTI-SENSE: N  
; ORIGINAL SOURCE:  
; ORGANISM: Plasmodium yoelii  
; STRAIN: 17X(NL)  
; DEVELOPMENTAL STAGE: erythrocytic stage  
; TISSUE TYPE: Blood  
; CELL TYPE: erythrocytic stage  
; IMMEDIATE SOURCE:  
; LIBRARY: Py-lambda gtl1-2-7 kb genomic expression

```

; CLONE: Py10.1111
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 718..3195
; OTHER INFORMATION:
US-07-638-431-1

```

	Query Match	2.7%	Score 63.2	DB 1	Length 4673
	Best Local Similarity 54.2%	Pred. No. 6.5e-07			
	Matches 128	Conservative 0	Mismatches 108	Indels 0	Gaps 0
Qy	2061	attaaacccgcatgatcctgctacaccaaatagtctgaaacgcgaactacaccagaac	2120		
Db	1626	AGTAATCTCAACAGATCAAAACGATCCAAACACCCCAACCAACCAATATACCCCAACAA	1685		
Qy	2121	ccctggagacacctaatctcccaaaaacacaaagactcctgaaatcctgggacacctca	2180		
Db	1686	CCCAAAACCCCAANTATACCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCA	1745		
Qy	2181	aactcctaatacacctataactcgcggaaattcctttlaactccgaagaacgcctaagaaccc	2240		
Db	1746	CAATCTCAAAATTAACCCAAATAACCCCAACCAACCAACCAATATACCCCAANTATACCCCAACCAACCC	1805		
Qy	2241	tgaaccccaaaactaataatcgttgcccacaaactgggaaataatgcgaataaagcca	2296		
Db	1806	AAATAACCCCAACCAACCAANTATACCCCAANTATACCCCAANTATACCCCAANTATACCCCA	1861		

Search completed: September 22, 2002, 05:22:06  
Job time: 20661 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2002, 01:51:46 ; Search time 1031.3 Seconds  
(without alignments)  
3940.596 Million cell updates/sec

Title: US-09-995-587A-2  
Perfect score: 2367  
Sequence: 1 atgtataaaacggtaaaaa.....ttaacaaacgtcgatttaac 2367

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457231 residues  
Total number of hits satisfying chosen parameters: 3472872

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```

Database :
1: /SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
C	1	80	3.4	10317	21	AAZ5988	Vector pGenDell seq
	2	78.4	3.3	1668	21	AAZ57329	SacB and cpy chimera
	3	78.4	3.3	1722	21	AAZ57330	SacB and cpy chimera
	4	71.8	3.0	2537	23	AAS82323	DNA encoding novel
	5	71.8	3.0	2787	23	AAS77631	DNA encoding novel
	6	71.8	3.0	2787	23	AAS86008	DNA encoding novel
	7	71.8	3.0	2787	23	AAS86581	DNA encoding novel
	8	71.8	3.0	2787	23	AAS88736	DNA encoding novel
	9	71.8	3.0	2787	23	AAS91108	DNA encoding novel

10	70.2	3.0	2079	23	AA585786	DNA encoding novel
c 11	70.2	3.0	2079	23	AA585998	DNA encoding novel
12	70.2	3.0	2079	23	AA587342	DNA encoding novel
c 13	70.2	3.0	2079	23	AA588703	DNA encoding novel
14	69	2.9	2325	23	AA569777	DNA encoding novel
15	69	2.9	2325	23	AA575941	DNA encoding novel
16	69	2.9	3084	23	AA585932	DNA encoding novel
c 17	69	2.9	3084	23	AA587383	DNA encoding novel
18	68.8	2.9	3730	17	AAT03195	Group B Streptococ
19	68.2	2.9	2745	23	AA573811	DNA encoding novel
20	68.2	2.9	5912	23	AA586048	DNA encoding novel
21	67.8	2.9	1690	23	AA585792	DNA encoding novel
22	67.8	2.9	1690	23	AA586019	DNA encoding novel
23	67.8	2.9	1690	23	AA587350	DNA encoding novel
c 24	67.8	2.9	3288	23	AA587395	DNA encoding novel
25	67.8	2.9	3288	23	AA585789	DNA encoding novel
26	67.8	2.9	3288	23	AA586010	DNA encoding novel
c 27	67.8	2.9	3288	23	AA587346	DNA encoding novel
c 28	67.8	2.9	3288	23	AA588746	DNA encoding novel
29	67.4	2.8	2568	23	AA568687	Mutant levansucras
30	67.4	2.8	3305	18	AAT68806	Purinegenic recepto
31	67.4	2.8	6721	24	AA518599	Purinegenic recepto
32	67.4	2.8	6721	24	AA518600	Purinegenic recepto
33	67.2	2.8	1584	23	AA568494	DNA encoding novel
c 34	67.2	2.8	1629	23	AA585800	DNA encoding novel
c 35	67.2	2.8	1629	23	AA586022	DNA encoding novel
c 36	67.2	2.8	1629	23	AA587359	DNA encoding novel
37	67.2	2.8	1935	23	AA570417	DNA encoding novel
38	67.2	2.8	1974	23	AA585799	DNA encoding novel
39	67.2	2.8	1974	23	AA586021	DNA encoding novel
40	67.2	2.8	2044	23	AA585834	DNA encoding novel
c 41	67.2	2.8	2044	23	AA587384	DNA encoding novel
42	67.2	2.8	2328	23	AA564228	DNA encoding novel
43	67.2	2.8	2361	23	AA564444	DNA encoding novel
44	67.2	2.8	2361	23	AA585823	DNA encoding novel
45	67.2	2.8	2370	23	AA569133	DNA encoding novel

## ALIGNMENTS

```

RESULT      1
AAZ35988/c
ID  AAZ35988 standard; DNA; 10317 BP.
XX
XX  AC  AAZ35988;
XX
XX  DT  09-FEB-2000 (first entry)
XX
XX  DE  Vector pGendell sequence SEQ ID NO:1.
XX
XX  KW  DNA sequencing vector; nested deletion; transposition; contig;
XX  KW  truncated lacZ; origin of replication; pGendell; mapping; ss;
XX

```



Query Match 3.3%; Score 78.4; DB 21; Length 1668;  
Best Local Similarity 47.6%; Pred. No. 3.6e-08;  
Matches 463; Conservative 0; Mismatches 461; Indels 48; Gaps 6;

QY 983 cacaagaatggtcaggatcagctgttttgaacagtgataactctctatccaatattttata 1042  
|||||  
Db 725 cacaagaatggtcaggatcagctgttttgaacagtgataactctctatccaatattttata 784  
|||||

QY 1043 caagggtagacagctgtgatacaataaccataatccataaataattgtagcgtactcttt 1102  
|||||  
Db 785 c-----tgatttctcggtaaacattacggtgcaaacacactgacacactgcaagaagta 838  
|||||

QY 1103 attaaactgataataatggaaatgatacactcgtcaggtacgaaatgactatattgtat 1162  
|||||  
Db 839 acgtatcagatcagacagctcttgaacatcaacggtgtagaggattataaatcaatct 898  
|||||

QY 1163 ttgaaggtgatgctattactaccacaaacttatgatcaatgga---aagctactaacaaga 1219  
|||||  
Db 899 ttgacggtgacggaacacgttacaaaattacacagcttcacgtatcagtagaagccaactaca 958  
|||||

QY 1220 gtccgataatattgaatcgtgtgctcatgttaattgaagatggttaattgtagcgtt 1279  
|||||  
Db 959 gtcacggcgaacaccatcagctgagagatctcactacgttagaagataaaggccacaat 1018  
|||||

QY 1280 acctgttttgaacagactggttttgaaaattatcaagcgagagaccacaaatttata 1339  
|||||  
Db 1019 acttagtatttgaacacacactggaactggaatggtaccgaagcggaagaatctttat 1078  
|||||

QY 1340 actggttaattatggtcggagatgacgcatatttaataacagagcttatttagaattcttt 1399  
|||||  
Db 1079 ttaacaaagcactactggaacaaagcacatcattcttccgtcgaagaagccaacacctc 1138  
|||||

QY 1400 ccaatgatgataatgaagtcgggcaacttgggttaattgagctatcgttactcacaac 1459  
|||||  
Db 1139 tgcaagcgataaaacacgcacgctgagttagcaaacggtcctcgtgtatgatgagc 1198  
|||||

QY 1460 taaataagacgcaagaatcctaaagtgagcagaggttatactcaccataatttctgcac 1519  
|||||  
Db 1199 taaacagtgattacacactgaaaaagtgatgaa-----accgctgattgcatcta 1249  
|||||

QY 1520 caatgtagcagtgataatgagcgcaacaaatgtagttaaattaggttaataaataattact 1579  
|||||  
Db 1250 acacagtaacagatgaatttgaacgcgcaacgtctttaaataatgaacgcaaatggtacc 1309  
|||||

QY 1580 tatttcgcgtaccgttttaaatcgaggggaagtaattgatgatgcttggatgaatgctaatt 1639  
|||||  
Db 1310 tgttactgactcccgatcaaaaatgacgattgacggtacgttacttaacgatattt 1369  
|||||

QY 1640 atgctgtgtgataatgttgaatggtcggatattgtgtgatgtacttaactggatctt 1699  
|||||  
Db 1370 a-----catgctgtgtatttttctaatctttaaactggcccat 1408  
|||||

QY 1700 ataagccataaattgattgtagtatttctgactgtcttctgttccctgcaaacctggcgga 1759  
|||||  
Db 1409 acaagcgtgacaaaactgcttctgtgttaaaaatgattctgtatccttaacagatgtaa 1468  
|||||

QY 1760 cagcaacttattcattatgtgtgtcccggttgcggaaagatgaccagaattattgatta 1819  
|||||  
Db 1469 ccttactcaccacacttctgtgtacc---tcaagcgaaggaacaaatgctgtgatta 1525  
|||||

QY 1820 cttcatatgactaatagaaatggagtagcgggttaaaggaatgattcacttggcgac 1879  
|||||  
Db 1526 caagctatatgaaacagagagattctacgcagacaaa-----caatcaacgtttgctgc 1579  
|||||

QY 1880 cgagtttcttactacaaataaacccggataacacaaactactgtttttagtcaaaatgacta 1939  
|||||  
Db 1580 caagcttctgtgtaacatcaaggcaagaaacacatctgttgtcaaaagacagcatccttg 1639  
|||||

QY 1940 atcaaggggatt 1951  
|||||  
Db 1640 aacaaggacaat 1651  
|||||

## RESULT 3

AAZ57330  
ID AAZ57330 standard; DNA; 1722 BP.  
XX  
AC AAZ57330;  
XX  
DT 03-APR-2000 (first entry)  
XX  
DE SacB and cpy chimeric gene #2.  
XX  
KW Levansucrase; SacB; cpy; salt tolerance; vacuole guide peptide;  
KW carboxypeptidase A; chimeric gene; transgenic plant; yeast;  
KW Bacillus subtilis 168; Saccharomyces cerevisiae X8;  
KW drought resistance; ds.  
XX  
OS Chimeric - Saccharomyces cerevisiae.  
OS Chimeric - Bacillus subtilis.  
XX  
PN CN1231337-A.  
XX  
PD 13-OCT-1999.  
XX  
PF 08-APR-1998; 98CN-0101336.  
XX  
PR 08-APR-1998; 98CN-0101336.  
XX  
PA (GENE-) INST GENETICS CHINESE ACAD SCI.  
XX  
PI Chen S, Zhang H, Dong W;  
XX  
DR WPI; 2000-087902/08.

Method for improving plant salt resistance using transgenic technology  
Example; Fig 4; 12pp; Chinese.

A method has been developed for raising salt tolerance of plants using transgenic technology. The method comprises: cloning levansucrase gene SacB and vacuole guide peptide (cpy) sequence of carboxypeptidase A from Bacillus and DNA of yeast, using the two kinds of gene to create a chimeric gene, and using the chimeric gene to structure plant expression carrier, using the obtained plant expression carrier to transform a plant and screen for resistant seedlings. The method can obtain drought-resistant salt-tolerance plants, and the polymerase chain reaction (PCR) and Northern analysis of transgenic plants shows that the exogenous gene is integrated in the salt-tolerance transgenic plant. The present sequence represents a SacB and cpy chimeric gene from the present invention.

Sequence 1722 BP; 590 A; 390 C; 339 G; 403 T; 0 other;

## Query Match

Best Local Similarity 3.3%; Score 78.4; DB 21; Length 1722;  
Matches 463; Conservative 0; Mismatches 461; Indels 48; Gaps 6;

QY 983 cacaagaatggtcaggatcagctgttttgaacagtgataactctctatccaatattttata 1042  
|||||  
Db 756 cacaagaatggtcaggatcagctgttttgaacagtgataactctctatccaatattttata 815  
|||||

QY 1043 caagggtagacagctgtgatacaataaccataatcaaaaaattgtagcgtactcttt 1102  
|||||

Db 816 c-----tgatttctcggtaaacattacggtgcaaacacactgacacactgcaagaagta 869  
|||||

QY 1103 attaaactgataataatggaaatgatacactcgtcgtcgggtacgaaatgactatattgtat 1162  
|||||

Db 870 acgtatcagatcagacagctcttgaacatcaacggtgtagaggattataaatcaatct 929  
|||||

QY 1163 ttgaaggtgatgctattactaccacaaacttatgatcaatgga---aagctactaacaaga 1219  
|||||

Db 930 ttgacggtgacggaacaaacttatacaaatgtacagcggttcacgtatcagtagaagccaactaca 989  
|||||





Db 1426 acaagcgctactacgcgccgacgaactcttccgtataaagaaagccagaagcttcagc 1485  
Qy 1403 atgatgatattaaagtcggaactctgggtaataatgcagctatcggtatctccaactaa 1462  
Db 1486 agagcgtataaagcgcgatctgagttagcgaacgcgccctcggtatcatagagttaa 1545  
Qy 1463 ataaggacgaagaatacctaaagtggtgagagttatactaccattatcttcgaccaa 1522  
Db 1546 ataatgatcacattgaaagaagtaagag-----ccgctgatcactctcaaca 1596  
Qy 1523 tggtaagcgataaattgagcgaccaaagtagtttaataataggtaataataattacttat 1582  
Db 1597 cggtaactgatgaatcagcgcgcaatgtttcaataaagaacggcaaatggtactgt 1656  
Qy 1583 ttgcgcgtaccggtttaaatcaggaagaatgatgatgcttgatgataatgataattat 1642  
Db 1657 tcaatgattcgcggtttcaaaaatgacatcgatcgatggtattaaactcaacgatattta-- 1714  
Qy 1643 ccggttgatgaattgtgcaatgctcgatattgtgctgatgtagtctaaactgatcttata 1702  
Db 1715 -----catgctggctgattatcatcaaaactctttaaccggcccttaca 1755  
Qy 1703 agcattaaatgattcgagtagtcttgactgcttcttctcgtcaaaactggcgacag 1762  
Db 1756 agcgcgtgaacaaacaggcgttctgctgcaaatgggtcttgatccaaacgatgtgacat 1815  
Qy 1763 caattattcatatgtctgccccgttgcgcgaagaagatgacaaagatattagtaact 1822  
Db 1816 tcaattactcactctcagtcagtcgcaagc---aaaggcaacaatgtggtattacaaa 1872  
Qy 1823 catatgatctaataagaaatggatgtagcgggttaaggaaatggattcaacttgggcacga 1882  
Db 1873 gctcatgacaaacagaggtcttcttcgaggataaa-----aaggcaacatttgcgcaa 1926  
Qy 1883 gttcttactacaaataaaccgcatgaacacaaactactgttttagctaaatgactaatc 1942  
Db 1927 gttcttaatgaacatacaaaaggcaataaaacatccgttgcataaaacagatcctcggagc 1986  
Qy 1943 aaggggattgatttgg 1959  
Db 1987 aaggacagctgacatgg 2003

## RESULT 5

ID AAS77631 standard; cDNA; 2787 BP.  
XX

AC AAS77631;  
XX

DT 13-FEB-2002 (first entry)  
XX

XX DNA encoding novel human diagnostic protein #13435.  
XX

Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX

OS Homo sapiens;  
XX

XX WO200175067-A2.  
PN

XX 11-OCT-2001.  
PD

XX 30-MAR-2001; 2001WO-US08631.  
PF

XX 31-MAR-2000; 2000US-0540217.  
PR

XX 23-AUG-2000; 2000US-0649167.  
PR

XX (HYSE-) HYSEQ INC.  
PA

XX Drmanac RT, Liu C, Tang YT;  
PI

XX WPI: 2001-639362/73.  
DR

XX P-PSDB; ABG13444.  
DR

XX

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX

PS Claim 1; SEQ ID No 13435; 103pp; English.  
XX

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SQ Sequence 2787 BP; 935 A; 596 C; 607 G; 649 T; 0 other;

## Query Match

Best Local Similarity 3.0%; Score 71.8; DB 23; Length 2787;

Matches 453; Conservative 0; Mismatches 482; Indels 42; Gaps 5;

Qy 983 cacaagaatggtcaggatcagctgttttgaacagtgataactctatccaatttttata 1042  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 365 cgcaagaatggtccggttctgcacaccttaccatgtgacgaaataccgtttattctaca 424  
Qy 1043 caagggtagacagctctgatacaaataccaatcatcaaaaattgctagctactcttt 1102  
Db | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | |||||

Db 425 ctgactattccggttaaacattacgccaacaaagcctgcaacagcgaggttaattgt 484  
Qy 1103 atttaactgataataatggaaatgtatcactcgtcaggtcagcaaatgactattgtat 1162  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 485 taaaatctgatgacacactcaaaatcaacggagtggaagatcaaaagcagattttgacg 544  
Qy 1163 ttgaaggtgatggctattactacaaacttatgatcaatggaaagctactacaagaagt 1222  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 545 gagcggaaaaacatacatcagaacgtttcagcagcttatcgatgaaggcaatatacatccg 604  
Qy 1223 ccgataatttcgaatcgctgctgcatgtaattgaagatgtaattgctgctgctgctacc 1282  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 605 gcgcacaccatcagctgagagaccctcactcagttgaaga---caaaagccataaaatcc 661  
Qy 1283 ttgtttttgaagcaagtactgttttgaaaaattatcaaggcaggaccataattataact 1342  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 662 ttgtattcgaagcccaacacgggaacaggaacggataccaagggcaggaattcttatta 721  
Qy 1343 ggttaaatattggcgagatgacgcatttaataatcaagagcttatttgaattcttcca 1402  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 722 acaaaagcgtactacgcgcgccgacgaactcttcgtaaaagaaagcagaagcttcagc 781  
Qy 1403 atgatgatattaaagtcggtggaactctgggtaattgagctatcggtatctccaactaa 1462  
Db | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | |||||

Db 782 agagcgtataaaacacgcatgctgagtagcgaacggtccctcgttatcatagagttaa 841  
Qy 1463 ataaagacgaagaaatccttaaggtgagagattactaccatttaatttctgcacaa 1522  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 842 ataagtattcacacattgaaaaagtaataag-----ccgctgactcactcaaca 892

QY 1523 tggtaagcagatgaattgagcgaccacaaatgtagttaaattaggttaataataattacttat 1582  
 Db 893 cggtaactgatgaatcgagcgcggaatgttttcaaatgaacgcgaatggtacttgt 952  
 QY 1583 ttgcgctaccggtttaaactcagggaggaagtaatgatgcttggatgaatgtaattatg 1642  
 Db 953 tcactgattcacgcggttcaaaaatgacgatcgatggtattaaactcaaacgatattta-- 1010  
 QY 1643 ccgttgatgataatgttgcgaatggtcggtatgttgcctgatgacttaactggtatcttata 1702  
 Db 1011 -----catgcttggttatgtatcaaaactcttaacgcgcctttaca 1051  
 QY 1703 agccattaaatgattcgtgagtagtcttgcctgcttcttgcctcaaaactgcgacag 1762  
 Db 1052 agccgctgaacaaacagggctgtgctgcaaatggttctgatacacaacgatgtgacat 1111  
 QY 1763 caacttattcattatgtctgtcccggttgcgggaagaatgacccaagtattgtactt 1822  
 Db 1112 tcacttactctcacttcgcagtcgcgcaagcc---aaaggcaacaatgtggttatcaca 1168  
 QY 1823 catatatgactaatagaatggagtagcggttaaaaggaatgattcaacttgggcaccga 1882  
 Db 1169 gctacatgaacaacagaggtctcttcgaggtacaa-----aaggcaacatttgcgcca 1222  
 QY 1883 gtttcttactacaaataaccgcggataacacacactactgttttttagctaaatgactaaac 1942  
 Db 1223 gcttcttaatgaacatacaaggaacataaaacatccgttgcataaaacacgcatcctggagc 1282  
 QY 1943 aaggggattggttg 1959  
 Db 1283 aaggacagctgacatgg 1299

## RESULT 6

AAS86008  
 ID AAS86008 standard; cDNA; 2787 BP.

AC AAS86008;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #21812.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 QS Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-P5DB; ABG21821.

XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity

XX Claim 1; SEQ ID No 21812; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostic products as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations in  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 2787 BP; 935 A; 596 C; 607 G; 649 T; 0 other;

Query Match 3.0%; Score 71.8; DB 23; Length 2787;

Best Local Similarity 46.4%; Pred. No. 1.3e-06;

Matches 453; Conservative 0; Mismatches 482; Indels 42; Gaps 5;

QY 983 cacaagaatggtcagatcagctgttttgacagtgataactctctatccattttata 1042  
 Db 365 cgcaagaatggtcgggttctgcaacctttacatgtgacggaataatcgttttatctaca 424

QY 1043 caagggtagacacgctgtgataacaataaccatacatcaaaaattgtagcgctactcttt 1102  
 Db 425 ctgactattcttgtaaacattacgccaacaaagcctgacacagcgaggttaattgt 484

QY 1103 atttaactgataataatggaattgtatcactcgtcaggtacgaaatgactattgtat 1162  
 Db 485 taaatctgatgacacactcaaaatacaacggagtggaagatcacaaagcgattttgacg 544

QY 1163 ttgaaggtgatgctattactaccaaacttatgatcaatggaagctactaaacaaaggtg 1222  
 Db 545 gagcggaaaacacataatcagaacgttcagcagcttatcgtgaaggcaattatcacatccg 604

QY 1223 ccgataatttgcaatgcgtgatcgtcatgtaattgaagatggttaattgtagcggtacc 1282  
 Db 605 gcgacacaccatacgtgagagacccctcacctacgttggaaga---caaaggccataaatacc 661

QY 1283 ttgtttttgaagcaagtactgtgttgaaaaattatcaaggcgaggacaaatttataact 1342  
 Db 662 ttgtattcgaagccaacacgagggaacagaaacggataccaaagcgagaagtctttatta 721

QY 1343 ggttaattatggcgagatgacgcatttataatcaagagcttatttagaattctttcca 1402  
 Db 722 acaagcgtactacgcgcgcgacgaactcttcctgtaaaagaaagccagaagcttcagc 781

QY 1403 atgatgatattgaagatcgggcaacttggcctaatacagcatcgttatcctcaaacataa 1462  
 Db 782 agagcgtaaaaaacgcgatgctgagtagcgaacggtccctcgttatcatagagttaa 841

QY 1463 ataaggacgaaaaagaatccttaaggtggcaggttatcaccatttaattcttcgaccaa 1522  
 Db 842 ataattgattacacattgaaaaaagttaag-----ccgctgatacattcaaca 892

QY 1523 ttgtaagcagatgaattgagcgacacaaatgtagttaaattaggttaataataattacttat 1582  
 Db 893 cggtaactgatgaatcgagcgcgcaatgttttcaaaatgaacgcgcaaatggtacttgt 952

QY 1583 ttgcgctaccggtttaaatcgcaggaagtaattgatgactgcttggaatgcaattatg 1642  
 Db 953 tcactgattcacgcggttcaaaaatgacgatcgatggttatcaactcaacgacattta-- 1010

QY 1643 ccgttggtgataatgttgcaatgggtcggtatgttctgtagacttaactggatcttata 1702  
 Db 1010

Db 1011 -----catgctggttatgtatatacaaaactctttaacgcgccttaca 1051  
 Oy 1703 agccattaaatgattcgtgagtagtctgtctctgttcctgcacaaactgcgcgacag 1762  
 Db 1052 agccgtgcaaaaacagggctgtgctgcaaatgggtcttgatccaacgagtgtgacat 1111  
 Oy 1763 caactattcatattatgctcccggtgcgcggaagaaagatgacaaactattgattctt 1822  
 Db 1112 tcacttactctacttcgcaagtcgcgaagcc---aaagcacaatgtggttatacaaa 1168  
 Oy 1823 catatgactaatagaaatgagtagcgggttaaaggaaatggattcaactctgggcacga 1882  
 Db 1169 gctacatgcaaacagaggtctcttcgaggataaa-----aagcacaattgcgcgcaa 1222  
 Oy 1883 gttcttactacaaataaccggtataacacaaactactgtttttagtctaaatgactaatc 1942  
 Db 1223 gcttctaatgacatcaagcaataaaacatccgttgtcaaaaacagcatccttgagc 1282  
 Oy 1943 aaggggattggattgg 1959  
 Db 1283 aaggacagctgacatgg 1299

## RESULT 7

AAS86581

ID AAS86581 standard; cDNA; 2787 BP.

AC AAS86581;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #22385.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN W0200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG22394.

XX New isolated polynucleotide and encoded polypeptides, useful in

XX diagnostics, forensics, gene mapping, identification of mutations

XX responsible for genetic disorders or other traits and to assess

XX biodiversity.

XX Claim 1; SEQ ID No 22385; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

XX polypeptide (II) sequences. (I) is useful as hybridisation probes,

XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome

XX and gene mapping, and in recombinant production of (II). The

XX polynucleotides are also used in diagnostics as expressed sequence tags

XX for identifying expressed genes. (I) is useful in gene therapy techniques

XX to restore normal activity of (II) or to treat disease states involving

XX (II). (II) is useful for generating antibodies against it, detecting or

XX quantitating a polypeptide in tissue, as molecular weight markers and as

XX a food supplement. (II) and its binding partners are useful in medical

XX imaging of sites expressing (II). (I) and (II) are useful for treating

XX disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 2787 BP; 935 A; 596 C; 607 G; 649 T; 0 other;

## Query Match

Best Local Similarity 3.0%; Score 71.8; DB 23; Length 2787;

Matches 453; Conservative 0; Mismatches 482; Indels 42; Gaps 5;

Oy 983 cacaagaatggtcaggtacagctgttttgaacagtgataaactctatcattttttata 1042

Db 365 cgaagaatgggtccggttctgcaacctttacatgtgcggaataatcctttatttaca 424

Oy 1043 caagggtagcacgctctgatacaataccaatcatcaaaaaattgctagcgtactctt 1102

Db 425 ctgactattccggtaaacattacggtcaaaaacagcctgcacacagcgaggtaatgt 484

Oy 1103 atttaactgataataatggaatgtatcaactcgtcaggtacgaaatgactatattgt 1162

Db 485 taaatctgatgacacactcaaatcaacggagtggaagatcaacaagcagattttgacg 544

Oy 1163 ttgaaggtgatggtctattactacaaacttatgatacaatggaagctactacaagaagt 1222

Db 545 gagacgaaaaacatatacgaacgttccagcagcttatcgatgaaggcaattatcacatcg 604

Oy 1223 cagataattgcaatgcgtgtagtgcctgtaattgaagatggtgaatggtgacgtacc 1282

Db 605 gcgacaacacatacgtgagagacccctcactacgttgaga---caaagccatacaatcc 661

Oy 1283 ttgttttgaagcaagtactggttttgaaaaattatcaaggcaggaaccataattataact 1342

Db 662 ttgtattcgaaacacacacgggaacagaaaacggatcaccaaggcgaagaattttatta 721

Oy 1343 ggttaaatctgcggagatgacgcatttaatacaagagcttatttgaattcttcca 1402

Db 722 acaagcgtactacgcggcgccgacgaacttcttcgtaaaagaaagcagaagcttcagc 781

Oy 1403 atgatgataataagagtcggcgaacttggtggttaataatgcagctatcggtatcccaactaa 1462

Db 782 agagcgtaaaaaacgcgatgctgagttagcgaacgcgcctcgttatcatagagttaa 841

Oy 1463 ataggcagaaaaaatctcaagggtgcagagttatactaccatttaattctgcaccaa 1522

Db 842 ataatgattacacattgaaaaaagtaataag---ccgctgatacttcaaca 892

Oy 1523 tggtaagcagataaattgagcgaacaaatgtagttaaattaggttaataataattacttat 1582

Db 893 cggtaactgataaactgcgcgcgaatgttttcaaaatgaacggaatggtactgt 952

Oy 1583 ttgcccgtacccttttaaatcgaggaaatgagatgagctgtggtgagatgaatgactat 1642

Db 953 tcaactgattcacgcggttcaaaaatgcgacatcgatggtatttaactcaaacgatttata 1010

Oy 1643 ccggttggtataatgttgcaatggtcggatattgttgcgtatgcttaactgagatttata 1702

Db 1011 -----catgcttggttatgtatatacaaaactctttaacgcgccttaca 1051

Oy 1703 agccattaaatgattcgtgagtagtctgtgactgcttctgttctctgcacaaactgcgcgacag 1762

Db 1052 agcgcgtcaaaaacagggctgtgctgcaaatgggtcttgatccaacgagtgtgacat 1111

Oy 1763 caacttattcatattatgctgtcccggtgcgcggaagaaagatgacaaagattagttactt 1822

Db 1112 tcacttactctcactgcagtcgcgaagcc---aaaggcaacaatgtggttatacaaa 1168



Db 1283 aaggacagctgacatgg 1299

RESULT 9  
 AAS91108  
 ID AAS91108 standard; cDNA; 2787 BP.  
 XX  
 AC AAS91108;  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE DNA encoding novel human diagnostic protein #26912.  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSER INC.  
 XX  
 PI Dmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI: 2001-639362/73.  
 DR P-PSDB; ABG26921.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 1; SEQ ID NO 26912; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 2787 BP; 935 A; 596 C; 607 G; 649 T; 0 other;

Query Match  
 Best Local Similarity 3.0%; Score 71.8; DB 23; Length 2787;  
 Matches 453; Conservative 0; Mismatches 482; Indels 42; Gaps 5;  
 Qy 983 cacaagaatggtcaggatcagctgtttgacagtgataactctatccaattttata 1042  
 Db 365 cgcagaatggttcggttctcgaaacctttacatgtgcggaaaaatccgtttattaca 424

Qy 1043 caagggtagacacgtctgtataacaataccaatacaataaaattgtcagcgtactcttt 1102  
 Db 425 ctgactattccggttaaacatttacggcaaacaaagcctgacacagcgaggtaaatgct 484  
 Qy 1103 atttaacigataaataatgaaatgtatcactcgtcaggtacgaaatgactatatgtat 1162  
 Db 485 taaaatctgtagacacactcaaatcaacgagtggaagatcacaaacgcatgtttgacg 544  
 Qy 1163 ttgaaggtgtaggtctattactaccacaaacttatgataatgaaagctactaacaaagtg 1222  
 Db 545 gagacggaataacataatcagaacgttcagcagcttatcgatgaaggcaattatcatccg 604  
 Qy 1223 ccgataatattgcaatgcgtgatctcatgtaattgaagatggttaaagtgatcggtacc 1282  
 Db 605 gcgacaaccaacgctgagagacctactacgttggaaga---caaagggccataataacc 661  
 Qy 1283 ttgttttgaagcaagtactgtttgaaaaattatcaaggcgaggacacaaatttataact 1342  
 Db 662 ttgtattcgaagcaaacacaggggaacagaaacgataccaaaggcggaatctttatta 721  
 Qy 1343 ggttaaatattgcggagatgacgcatttaataatcaagagcgtatttagaattctttcca 1402  
 Db 722 acaaacgctactacggcgcgacgaactcttcctgtaaaagaaagcagaagcttcagc 781  
 Qy 1403 atgatgattataagatcggtggcaacttggttaagcagctatcggtatctcctcaactaa 1462  
 Db 782 agagcgtaaaaaacgcgagtctgagttagcgaacggtccctcggtatcatagagttaa 841  
 Qy 1463 ataaggacgaaaagaatcctaaagtggtgagagttatactcaccatttaattctgcacaa 1522  
 Db 842 ataattgattacacatigaaaaaagtaagaag-----ccgcgtgactcactcaaca 892  
 Qy 1523 tgtaagcagatgaattgagcgaccacaaatgtagttaaattaggtataataataattacttat 1582  
 Db 893 cgtaactgatgaatcgagcgcggaatgttttcaaaatgaaacggcaaatgtagctgtg 952  
 Qy 1583 ttgcgcgtaccggttttaaatcgaggaagtaagtaagatgctgtggaatgtaataatg 1642  
 Db 953 tcaactgattcacgcggttcaaaaatgacatgctggtatttaactcaaacgatatta-- 1010  
 Qy 1643 cgtgtggtataatgttgcgaatggtcggatattgtctgtagatgctcaactggatctata 1702  
 Db 1011 -----catgctgtgtatgtatcaaaacttttaacgcgaccttaca 1051  
 Qy 1703 agccattaaatgattctgtagtcttgaactgtcttcttctcgtcaaacctggcgacag 1762  
 Db 1052 agccgtgaacaaaacagggcttctgctgcaaatgggtctgtatccaaacagatgacat 1111  
 Qy 1763 caactattcatattatgctgtcccggttgcggaaaagatgaccaaagattagttactt 1822  
 Db 1112 tcaacttctcactctcgcagtgccgcaagcc---aaaggcaacaatgtggttatacaca 1168  
 Qy 1823 catatgactaataagaatggagtagcgggttaaaggaaatgattcaacttgggcacga 1882  
 Db 1169 gctacatgacaaaacagaggtcttctcgaggataaa-----aaggcaacatttgcgcaa 1222  
 Qy 1883 gtttctactacaataataccgagataacacaaactactgttttagctaaaatgactaatc 1942  
 Db 1223 gcttttaagtgaacatcaaaaggcaataaataacatccgttgtcaaaaacagcatctcgtg 1282  
 Qy 1943 aagggtgattggtgg 1959  
 Db 1283 aaggacagctgacatgg 1299

RESULT 10  
 AAS85786  
 ID AAS85786 standard; cDNA; 2079 BP.  
 XX  
 AC AAS85786;  
 XX  
 DT 13-FEB-2002 (first entry)



[illegible]







AAS88703/c  
ID AAS88703 standard; cDNA; 2079 BP.  
XX  
AC AAS88703;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #24507.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR P-PSDB; ARG24516.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
PS Claim 1; SEQ ID NO 24507; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2079 BP; 461 A; 475 C; 433 G; 710 T; 0 other:

	Query Match	3.0%	Score	70.2	DB	23;	Length	2079;
	Best Local Similarity	46.9%;	pred.	No. 2.7e-08;				
	Matches	542; Conservative	0;	Mismatches	553;	Indels	60;	Gaps
Qy	817	ggacaagtgcgaattggaatggatccaactgtcatcgcaatgatgggaattccaaac	876					
Db	1833	GGACAGTAGCTGAATACAACGGCATCACGTTGTGTTCTCTTCGGGAAACCGGAAG	1774					
Qy	877	caaaatgataatcatactatccttataataaagtatgttgataatgaattaagaatcat	936					
Db	1773	ACGCTGATGACACATCAACTACATGTTTTATCAAAGCTGGAAAACCGGGCCGTGTC	1714					

Qy	937	tggaagaatgtagtccaaatttttgctcataattctacc-----gcggtttccaagaa	990
Db	1713	TTTAAAGACAGCGATAAGTTCGACGCCACGATCCGATCCTGMAAGATGACAGCGAAGAA	1654
Qy	991	tgtcaggatcaagtcttttgaaacagtgataaactctaccaattatttatacaaggta	1050
Db	1653	TGTCGCGGTTCTCAACCTTTACATCTGACGGAAATATCCGTTATTCTACAC-----T	1600
Qy	1051	gacagctctgataaacaataccaatacatcaaaaaatgtgacgctactcttttatttaact	1110
Db	1599	GACTATTCGGGTAAACATTTACGGCAACAACACCTGACAACAGCGCAGGTAATGTGTCA	1540
Qy	1111	gataataatggaaatgtatcacctgctcaggtacgaaatgactataattgtattgaaggt	1170
Db	1539	AAATCTGATGACACACTCTMAAATCAACGAGTGGAGATCACAACGATTTTTCACGGA	1480
Qy	1171	gatggctattactaccaaacattatgataatggga-----aagctactaacaagggtgcc	1224
Db	1479	GACGGAAACATATCAGAACGTTTCAGCAGTTTATCGATGAAGCAATTTATACATCCGC	1420
Qy	1225	gataatttgcgaatcgctgctcatgttaattgaagatgtaattggtgatcggtaccctt	1284
Db	1419	GACAACCATAGCTGAGAGACCCCTCACTACGTTGAAGA---CAAAGGCCATAAATACCTT	1363
Qy	1285	gtttttgaagcaagtactgctgtttggaaaaattcaaggcggaggaccaaaattataactgg	1344
Db	1362	GTATTCGAGGCCAACACCGGAACAGAAACGGATACCAAGCGCAAGCAATCTTTATTAC	1303
Qy	1345	ttaattatggcgagatgacgcatttaataatcaagcggaggaccaaaattataactgg	1404
Db	1302	AAAGCGTACTACGGCGCGCACGAATCTTCCTCGTAAAGAAAGCCAGAGTTCAGCAG	1243
Qy	1405	gatgatattaagagtcgggcaacttgggttaattgcagctatcggtatcctcaactaaat	1464
Db	1242	AGCGCTAAACACCGCATGCTGTGATTAGCGAACGGCGCCTCGGTATCATAGAGTTAAT	1183
Qy	1465	aaggacgaaagaatcctaaaggtggcagagttatactcaacttaattctgcaccaatg	1524
Db	1182	AATGATTACACATTGAANAAGTAATGAAG-----CGCTGATCACTTCAAAACAG	1132
Qy	1525	gtaagcgatgaattgacgcacaaatgatttaattaggtaaatgtaattgtaattatgcc	1644
Db	1131	GTAACTGATGAATAGCGCGCGCAATGTTTCAAAATGAACGGCAATGGTACTGTGTC	1072
Qy	1585	gcgcgtaccgctttaaactcggaagaatgatgatgcttggaatgtaattatgcc	1644
Db	1071	ACTGATTACCGGGTTCAAAAATGACGATCGATGGTATTAATCTCAACGATATTTA---	1014
Qy	1645	gttggtgataatgttgcaatggcggatattgtctgtagtagtctaaactggaatctataag	1704
Db	1015	-----CATGCTTGTTATGATATCAACTCTTTAACGGGCCCTTTACAGG	973
Qy	1705	ccattaaatgattctggagtagcttgactgctcttctcgtcaaaactggcgacaagca	1764
Db	972	CGCGTGACAAAACAGGCGTTGTGCTGCAAAATGGGTCTTGATCCAAACGATGTGACATC	913
Qy	1765	acttattcaattatgctgtccccggttcgcggaaaaagatgaccaagattagttacttca	1824
Db	912	ACTTACTCTCACTTCGAGTCCCGCAAGCC-----AAAGGCAACATGTGGTTATCACAAG	856
Qy	1825	tatgtactaatagaataatggatagcgggttaagggaatggattcaacttgggcaccaggt	1884
Db	855	TACATGACAAACACAGCGCTTCTTCGAGAGATAAAA-----AGCAACATTTGGCCAAAC	802
Qy	1885	ttcttactacaaataaacccggataaacaactactgttttagcttaaatgactaatcaa	1944
Db	801	TTCTTAATGAACATCAAAAGCAATAAANACATCCGTTGTCAAAACAGCATCTGTAGCAA	742
Qy	1945	ggggattggatttgg	1959
Db	741	GGACGCTGCATGG	727





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 21, 2002, 23:37:00 ; Search time 12398.7 Seconds  
(without alignments)  
3995.025 Million cell updates/sec

Title: US-09-995-587A-2

Perfect score: 2367  
Sequence: 1 atgtataaaagcgtaaaaa.....ttaacaaacgtcgatttaac 2367

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg\_hum.\*  
31: em.htg\_inv.\*  
32: em.htg\_other.\*  
33: em.htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
-----						

1	2367	100.0	2367	6	AX316614
2	2367	100.0	2394	6	AX316615
3	2367	100.0	2592	6	AX316616
4	718.6	30.4	4305	1	STRPTF
5	637	26.9	4634	6	AX316622
6	548	23.2	3600	1	STRPTF
7	97.8	4.1	11318	1	AE007686
8	84	3.5	155204	2	AC007926
9	80	3.4	2007	1	BSSACBG
10	80	3.4	5941	12	SYNPLA
11	80	3.4	6347	12	SYNPLA
12	80	3.4	6798	12	SYNPLA
13	80	3.4	10317	6	AR162197
14	80	3.4	88063	1	BS294043
15	80	3.4	209510	1	BSUB0018
16	78.4	3.3	5831	12	AF047518
17	78.4	3.3	5842	12	AF004910
18	78.4	3.3	5846	12	XXU17500
19	78.4	3.3	6349	12	AF047519
20	78.4	3.3	9297	12	AF048702
21	78	3.3	180668	2	AC020857
22	76.4	3.2	10115	5	AF397467
23	75.2	3.2	172307	2	AC044842
24	75	3.2	22398	5	PFU271723
25	75	3.2	164399	3	PFMAL3p6
26	73.6	3.1	1422	1	BSU34874
27	73	3.1	35793	5	AY016024
28	73	3.1	154918	9	AL138781
29	72.6	3.1	666	8	AF413050S2
30	72.2	3.1	393	1	MBQVSP02
31	72.2	3.1	3669	1	AF396972
32	72.2	3.1	3693	1	AF396971
33	72.2	3.1	16136	1	AF396969
34	72.2	3.1	16204	1	AF396970
35	72	3.0	586	1	AF362688
36	71.4	3.0	144505	2	AL662884
37	70.8	3.0	199551	2	AC006281
38	70.8	3.0	237258	2	AL606925
39	70.4	3.0	2167	3	PRSTARPA
40	69.6	2.9	53870	3	AE003163
41	69.4	2.9	185994	2	AC002042
42	68.8	2.9	532	1	AF362685
43	68.8	2.9	550	1	AF362687
44	68.8	2.9	586	1	AF362690
45	68.8	2.9	3730	6	I34014

## ALIGNMENTS

RESULT 1

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

Source

Location/Qualifiers

1..2367

/organism="Lactobacillus reuteri"

/db\_xref="taxon:1598"

BASE COUNT

ORIGIN

AX316614 Sequence

AX316615 Sequence

AX316616 Sequence

M18954 S.mutans fr

AX316622 Sequence

L08445 Streptococc

AE007686 Clostridi

AC007926 Trypanoso

X02730 Bacillus su

L05083 Cloning vec

L05082 Cloning vec

AR162197 Sequence

294043 B.subtilis

299121 Bacillus su

AF047518 Cloning v

AF004910 Cloning v

U17500 Cloning vec

AF047519 Cloning v

AF048702 Site-spec

AC020857 Mus muscu

AF397467 Ictalurus

AC044842 Homo sapi

AJ271723 Fugu rubr

Z98551 Plasmodium

U34874 Bacillus st

AY016024 Takifugu

AL138781 Human DNA

AF413051 Zea mays

AF162139 Mycoplasma

AF396972 Mycoplasma

AF396971 Mycoplasma

AF396969 Mycoplasma

AF396970 Mycoplasma

AF362688 Streptoco

AL662884 Homo sapi

AC006281 Plasmodiu

AL606925 Mus muscu

Z30339 P.reichenow

AB003163 Drosophil

AC002042 Homo sapi

AF362685 Streptoco

AF362687 Streptoco

AF362690 Streptoco

I34014 Sequence 8

AX316614

Sequence 2 from Patent WO0190319.

AX316614

AX316614.1 GI:17899728

Lactobacillus reuteri.

Lactobacillus reuteri.

Bacteria; Firmicutes;

Lactobacillus.

1 (sites)

van Geel-Schutten,G.H.,

Rahouli,H., Dijkhuizen,L. and van

Hijum,S.A.

Novel fructosyltransferases

Patent: WO 0190319-A 2 29-NOV-2001;

TNO (NL)

Location/Qualifiers

1..2367

/organism="Lactobacillus reuteri"

/db\_xref="taxon:1598"

814 a 433 c 459 g 661 t

Query Match		100.0%	Score 2367;	DB 6;	Length 2367;		
Best Local Similarity		100.0%;	Pred. No. 0;				
Matches 2367;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;		
QY	1	atgtataaaagcggtaaaatggcgagctgttacactctcgactgctgcgctgggtattt	60				
DB	1	ATGTATAAAAGCGGTAAAATGGGCAGTCGTTACACTCTCGACTGCTGCGCTGGTATTT	60				
QY	61	ggtcaacaactgttaaatgcctcgcggacacaaaatttgaacaaatgattcttctact	120				
DB	61	GGTCAACAACCTGTAATATGCATCCGGGACACAAATATTGAAAACAATGATTCTCTACT	120				
QY	121	gtacaagttacaacagtgataatgatattgtgtttaaagtgtagacacttggtagt	180				
DB	121	GTACAAGTTACACAGGTGATATGATATGTGTGTTAAAGGTGACACTGGTAGTGGT	180				
QY	181	caagttagtcagctgtagtacgactattagaactctctgctaagtcaaatagtgcttct	240				
DB	181	CAAGTTAGTCAGCTAGTGATACGACTATTAGAACTTCTGCTAATGCAATATAGTGTCT	240				
QY	241	tctgcgctaaatacacaaaattctaaacagtcgaagtctctgctgctgcaataacatca	300				
DB	241	TCTGCGCTAAATACACAAAATTTCTAACAGTCAAGTAGCAAGTTCTGCTGCAATAACATCA	300				
QY	301	tctacaagttccgcagcttcattaaataacacagatagtaaaagcgcctcaagaaaatact	360				
DB	301	TCTACAAGTTCCGCAGCTTCATTAAATAACACAGATAGTAAAGCGGCTCAAGAAAATACT	360				
QY	361	atacagccaaaatgatgacacgcaaaaagctgcaccagctaaacgaactctctcgaagct	420				
DB	361	AATACAGCCAAAATGATCACACGCAAAAAGCTGCACCAGCTAAACGAATCTTCTGAGCT	420				
QY	421	aaaaatgaaccagctgtaaacgttaaatgaattcttcagctgcgcaaaaatgatcatcaaaa	480				
DB	421	AAAAATGAACCAAGCTGTAACCGTTAATGATTTCTGAGCTGCAAAAATGATCATCAACA	480				
QY	481	tcacgttaaaagaatlaactaccctgaagttaaacaagagtgctgaaacggttgtaaaaaag	540				
DB	481	TCCAGTAAAAACAATACTACCCCTAAGTTAAACAAGGATGCTGAAAACCGTTGTA AAAAG	540				
QY	541	gcgggaattgatcctaacaagtttaactgatgaccagatataaagcattaaataagatgaac	600				
DB	541	GCGGGAATTTGATCCTAACAGTTTAACTGATGACCAGATTAAGCATTAAATAAGATGAAC	600				
QY	601	ttctcgaagctgcaaaagcttggttacacaaaatgactataatgatttccaaaagattgct	660				
DB	601	TTCTCGAAAGCTGCAAAAGTCTGGTACAAAATGACTTAATATGATTTCCAAAAGATTGCT	660				
QY	661	gatacgttaatacaacaagatggctcggtacacagttccattctttaaagcaagtgaaatc	720				
DB	661	GATAGTTTAATCAACAAGATGGTGGTACACAGTTCCTATTCTTAAACCAAGTGAATC	720				
QY	721	aaaaatatgcctgcgctacacaaatgaatgcaaaaactaatatctattgaaccttttagat	780				
DB	721	AAAAATATGCTTCGCCCTACAACTAAAAGATGCACAAACTTAATACATTATTGAACCTTTAGAT	780				
QY	781	gtatgggattcatggccagttcaagatgttcggacaggaagttgcttaattgggaatgc	840				
DB	781	GTATGGGATTCTATGCCAGTTCAAGATGCTCGGACAGGACAAAGTTGCTTAATTTGGAAATGGC	840				
QY	841	tatcaactgtcatcgaatgatgggaatttccaaacccaaaatgataatcatctatcttc	900				
DB	841	TATCAACTTTGTATCGCAATGATGGGAATTCCAAAACCAAAATGATAATCATATCTATCTC	900				
QY	901	ttataataaagttggtgataatgaattgaatcatgttaggaagaatgtaggtcccaattttt	960				
DB	901	TTATATAATAAGTATGGTGAATAAGTAATGATCAATTTGGAAGAAATGTAGTCCCAATTTT	960				
QY	961	ggctataaattctaccgcgggtttcaagaatgggtcagatcatcagctgttttgaacagtgat	1020				
DB	961	GCCTATAAATTCTACCGCGGTTTCACAAAGATGGTCAGGATCAGCTGTTTGNACAGTGAT	1020				

QY	1021	aactctatccaattattttatacaagggtagacacgctctgataaacaataaccaatcatcaa	1080				
DB	1021	AACCTATCCAATTTATTTATACAAAGGTAGACAGCTGTATAACAATACCAATCAATCAA	1080				
QY	1081	aaaattgctagcgtactcttatttaactgataataatggaaaatgtatcactcgtcag	1140				
DB	1081	AAAATGCTAGCGCTACTCTTTATTTAACTGATAAATGGAATGATATCATCTCGCTCAG	1140				
QY	1141	gtacgaataactactattgtatttgaagggtgtagtgcctattactaccaaaacttatgataa	1200				
DB	1141	GTACGAATAACTATATTGTATTGGAAGGTGATGGCTATTACTACCAAACTTTATGATCAA	1200				
QY	1201	tggaaagctactaacaagaagtgccgataaatttgaattgcaatggcgtgtagtcatgtaattgaa	1260				
DB	1201	TGGAAGCTACTAACAAAGGTGCCGATAATATTCANTGCTGATGCTCATGTAATTTGAA	1260				
QY	1261	gatggtaagtgtgacgtgtaacctgttttttgaagcaagtagtctgttggaaaaattatcaa	1320				
DB	1261	GATGGTAAATGGTGATCGGTACCTTGTGTTTGAAGCAAGTACTGTGTTGGAAAATTATCAA	1320				
QY	1321	ggcagagacacaaatttataactggttaaatatggcggagatgacgcatttaatatcaag	1380				
DB	1321	GGCAGAGACCAATTTATACTGGTTAAATATGCGCGAGATGACGCATTTAATATCAAG	1380				
QY	1381	agcttattagaattcttccaatgatgatttaagagtgctgggcaacttgggctaaatgca	1440				
DB	1381	AGCTTATTAGAAATTTCTTCCAAATGATGATATTAAGAGTCGGGCAACTTGGGCTAATGCA	1440				
QY	1441	gctatcggtatcctcaaaactaaataaaggacgaagaaatcctaagggtggcagagttatc	1500				
DB	1441	GCTATCGGTATCCTCAAACTAAAATAAGGACGAAAAGAAATCCTAAGGTGGCAGAGTTATAC	1500				
QY	1501	tcaccattaatttctgcaccaatggtaagcgatgaaattgagcaccacaaatgagttaa	1560				
DB	1501	TCACCAATTAATTTCTGCACCAATGGTAAGCGATGAATTTGAGCGACCAATTTAGTTAAA	1560				
QY	1561	ttaggttaataataattacttatttgcgcgtaccogtttaaatcgaggaagtaaatgatgat	1620				
DB	1561	TTAGGTAAATAAATATTACTTTATTGCGCTTACCCTTTAAATCGAGAAATTAATGATGAT	1620				
QY	1621	gcttgatgaatgctaatatgcgttggtgataatgtgcaatgttcgagatattgtct	1680				
DB	1621	GCTTGATGAATGCTAAATATGCGGTGGTGAATGTTCAATGTTCAATGGTTCGGATGTTGCT	1680				
QY	1681	gatagtcactggtgatttataagccattaaatgattcttgagtagtcttgactgctctct	1740				
DB	1681	GATAGTCTAACTGGATCTTATAAGCCATTAATGATTTCTGAGTAGTCTTTGACTGCTTCT	1740				
QY	1741	gttccctgaaactggcgacagcaacttattcattatattatgcttccccgttgcggaaaaa	1800				
DB	1741	GTTCCTGCAAACTGGCGGACAGCAACTTATTATATATTATGCTGTCCCGTTCGCCGAAAA	1800				
QY	1801	gatgaccaagttatagttacttctcatatatactgataaataagaaatggagtagcggtaaagga	1860				
DB	1801	GATGACCAAGTATTACTTCTCATATGACTAATAGAAATGAGTAGCGGTGAAGGA	1860				
QY	1861	atggattcaacttggcgacagcttcttactacaataatbaaccggataacacaaactact	1920				
DB	1861	ATGGATTCAACTTGGCACCCGAGTTTCTTACTACAAATTTAACCCGGATACACAACTACT	1920				
QY	1921	gttttagctaaaaatgactaatcaagggtatggatttggatgattcaagcgaataatctt	1980				
DB	1921	GTTTTAGCTAAAATGACTAATCAAGGGGATTTGGATTTGGATGATTCAAGCGAAAAATCTT	1980				
QY	1981	gatatgattggtgatttagactcccgctgttaccctggcgaaactgataaactgttgat	2040				
DB	1981	GATATGATTGGTGATTAGACTCCGCTGCTTTACCTGGCGAAGCTGATAAACCTGTTGAT	2040				
QY	2041	tgggacttaattgggttatggataaaacccgcatgacctgtctacaccaaataatgacctgaa	2100				
DB	2041	TGGGACTTAATTTGGTTATGGATTAAAACCGCATGATCTCTACACCAAATGATCTCTGAA	2100				

QY 2101 acgcaactacacagaaacccctgagacacctaataactccccaaacacccaaagactcct 2160  
Db 2101 ACGCCAATACACAGAAACCCCTGAGACACCTAACTACTCCAAAACACCAAGACTCCT 2160  
QY 2161 gaaaatcctgggacacccctcaaaactcctaatacacctaataactccggaattcctttaact 2220  
Db 2161 GAAAATCCTGGGACACCTCAAACTCCTAATACACCTAATACTTCCGGAATTCCTTTAACT 2220  
QY 2221 caagaaacgcctaagcaacctgaaacccaaactaataatcgtttgccaaactggaat 2280  
Db 2221 CCAGAAACGCCCTAAGCAACCTGAAACCCAACTAATAATCGTTGCCCAAACTGGAAT 2280  
QY 2281 aatgccataaagccatgatggcctagtagtggaacatctgcttagtatgttggctt 2340  
Db 2281 AATGCCATAAAGCCATGATGGCCTAGTATGGCAATCTGCTTAGTATGTTGGTCTT 2340  
QY 2341 gcagaaattaaacacgcctgatttaac 2367  
Db 2341 GCAGAAATTAAACACGTCGATTTAAC 2367  
RESULT 2  
LOCUS AX316615 2394 bp DNA linear PAT 14-DEC-2001  
DEFINITION Sequence 3 from Patent WO0190319.  
ACCESSION AX316615  
VERSION AX316615.1 GI:17899729  
KEYWORDS Lactobacillus reuteri.  
SOURCE Lactobacillus reuteri.  
ORGANISM Lactobacillus reuteri.  
REFERENCE 1 (sites)  
AUTHORS van Geel-Schutten,G.H., Rahaoui,H., Dijkhuizen,L. and van  
Hijum,S.A.  
TITLE Novel fructosyltransferases  
JOURNAL Patent: WO 0190319-A 3 29-NOV-2001;  
TWO (NL) Location/Qualifiers  
source  
1. .2394  
/organism="Lactobacillus reuteri"  
/db\_xref="taxon:1598"  
BASE COUNT 829 a 437 c 464 g 664 t  
ORIGIN  
Query Match 100.0%; Score 2367; DB 6; Length 2394;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2367; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 atgtataaaagcggtgaaataatggcgagtcgttacactctcgactgctgcgctgtattt 60  
Db 28 ATGTATAAAGCGGTAAATAATGGCGAGTCGTTACACTCTCGACTCTCGCTGCTGTAFTT 87  
QY 61 ggtgcaacaactgtaaatgcatccgagacacaaatattgaaacaaatgattcttctact 120  
Db 88 GGTGCAACAACGTAAATGCAATCCGCGGACACAAATATTGAAACAAATGATTCTTCTACT 147  
QY 121 gtacaagttaacaacagggtgaataatgatttgcgtttaaaggtgacacttggttagtgg 180  
Db 148 GTACAAGTTACAACAGGTGATAATGATATTGCTGTTAAAAGTGTGACACTTGGTAGTGGT 207  
QY 181 caagttagtgcagtagtagtacgactattgaaactctctgctaattgcaaatagtgtctt 240  
Db 208 CAAGTTAGTGCAGCTAGTGTATACACTATTAGAACTCTGCTAATGCAAAATAGTCTTCT 267  
QY 241 tctgcccttaatacacaaaattctaacagtcgaactagcaagttctgctgcaataacatca 300  
Db 268 TCTGCCGCTAATACACAAAATTCCTAACAGTCAAGTAGCAAGTCTGCTGCAATTAACATCA 327  
QY 301 tctacaagttccgcagcttctattataaataacacagatagtaaacgctcgaagaaataact 360  
Db 328 TCTACAAGTTCGCCAGCTTCTATTAAATAACACAGATAGTAAGCGGCTCAAGAAATACT 387

QY 361 aatacagccaaaatgatgacacgcaaaaaagctgcaccagctaacgaatctctgaagct 420  
Db 388 AATACAGCCAAAATGATGACACGCAAAAAGCTGCACAGTAACGAATCTTCTGAAGCT 447  
QY 421 aaaaatgaacacgctgtaaacgttaattgattcttccagctgcaaaaaatgatgatacaaa 480  
Db 448 AAAAAATGAACACGCTGTAAACGTTAATGATTCCTCAGCTGCAAAAATGATGATCAACAA 507  
QY 481 tccagttaaaaagaataactaccgcttaagttaaaaaaggatgctgaaaacgttgttaaaaaag 540  
Db 508 TCCAGTAAAAAAGAAATACTACCGCTAAGTTAAACAAGGATGCTGAAAAAGCTGTGTAAGAAAG 567  
QY 541 gcgggaattgattcctaacagtttaactgatgaccagattaaaagcattaaaataaagatgaac 600  
Db 568 GCGGGAATTGATCCTTAACAGTTTAACTGATGACAGATTAAAGCATTAATTAAGATGAAC 627  
QY 601 ttctgaaagctgcaaaagctggttacacaaatgacttaataatgatttccaaaagattgct 660  
Db 628 TTCTCGAAAGCTGCAAAAGCTGCTGATACAAAATGACTTTAATGATTTCCAAAAGATTGCT 687  
QY 661 gatacgttaatacaacaagaatggtcggtacacagttccattctttaaagaagtgaaatc 720  
Db 688 GATACGTTAATCAAAACAAGATGTCGCTACACAGTTCCATCTCTTTAAAGCAAGTGAATC 747  
QY 721 aaaaatagctgctccgtctacaactaaaagatgcacaaactaatactattgaaccttagat 780  
Db 748 AAAAAATAGCTGCTCCGCTACAACATAAGATGACAAACTAATTAATTAAGAACCTTTAGAT 807  
QY 781 gtatggattcatgycgagttcaagatgttcggacaggacaaagtgtgctaattggaatggc 840  
Db 808 GTATGGGATTCATGGCCAGTTCAAGATGTCGGACAGGACAAAGTTGCTAATTTGGAATGGC 867  
QY 841 tatcaactgtcatcgcaatgatgggaattccaaacaaataataataatcatatctatctc 900  
Db 868 TATCAACTTGTCTATCGCAATGATGGGAATTTCCAAACCAAAATGATAATCATATCATCTC 927  
QY 901 ttataataaagtatggtgataatgaattaaagtcatttgggaagaatgtaggccaattttt 960  
Db 928 TTATATAATAAGTATGCTGATATAATGAATTAAGTCATTTGGAAGAAATGTAGTCCAAATTTT 987  
QY 961 ggcataactctaccgcggtttcacaagaatggtcaggatcagctgttttgaacagtgat 1020  
Db 988 GGCATAATTTCTACCGCGGTTTCACAAGAATGCTCAGGATCAGCTGTTTGAACAGTGTAT 1047  
QY 1021 aactctatccaattatttatcaaggtgtagacacgtctgataacaataacaatcatcaa 1080  
Db 1048 AACTCTATCCAATTTTATACAAGGTGAGACACGCTCTGATAACAATACCAATCATCAA 1107  
QY 1081 aaaaatgctagcgcctactcttatttaactgataataatggaagaatgtatcactcgcctcag 1140  
Db 1108 AAAAAATGCTAGCGCTACTCTTTATTAACTGATAATAATGAAGAAATGTATCCTCCTCAG 1167  
QY 1141 gtacgaatgactatattgtattggaaggtgagtggttactaccacaaacttatgataa 1200  
Db 1168 GTACGAATGACTATATTGTTGAAGGTGATGGCTATTACTACCAAACTTATGATCAA 1227  
QY 1201 tggaaagctactaaacaaaggtccgataattgcaatggcgtgctgataatgaa 1260  
Db 1228 TGGAAAGCTTACTAAACAAGGTGCCGATAATATTGCAATGGCTGATGCTCATGTAATTGAA 1287  
QY 1261 gatggtaattggtgatcggtacacctgttttgaagcaagtcactggttttggaataattcaa 1320  
Db 1288 GATGTTAATGGTATCGGTACCTTGTTTTGAAGCAAGTACTGTTGGAAATATTACAA 1347  
QY 1321 ggcggaggaacaaatttataactgggttaaatattatggcggagatgacgcatttaataaag 1380  
Db 1348 GCGGAGGACCAAAATTTATAACTGGTTAAATATTATGGCGGAGATGACGCATTTAATATCAAG 1407  
QY 1381 agcttatttagaatctttcccaatgatattaaagtcgggcaacttgggctaataagca 1440  
Db 1408 AGCTTATTAGAAATCTTTTCCAATGATGATATTAAGAGTCGGGCAACTTGGGCTAATGCA 1467

```
QY 1441 gctatcggtatcctcaaaactaaataaagagcgaagaaatccttaagtgccagagttatc 1500
DB 1468 GCATCGGTATCCTCAAACTAAATAAGAGCAGAAAGAAAGTCCCTAAGTGGCGAGGTATAC 1527
QY 1501 tcacatttaattctgcacccaatgtaagcgtgaaatgagcgaccaaagttagttaa 1560
DB 1528 TCACCAATTAATTTCTGCACCAATGGTAAGCGATGAAATGGAGCAGCAAAATGAGTTAA 1587
QY 1561 ttaggtaataaattactatttcccgctaccggttttaaatcgagagaagaaatgat 1620
DB 1588 TTAGGTAATAATATATATTTGCGCGCTACCGGTTAAATCGAGGAAGTAATGATGAT 1647
QY 1621 gcttgatgaatgctaattatccgttggtgataatggttgcaatggtcgatatgttgc 1680
DB 1648 GCTTGGATGAATGCTAAATATATCGCGTGGTGATATGTTGCAATGTCGCGATATGTC 1707
QY 1681 gatagtcaactggatcttataaagccattaaatgattctggagtagtcttgactgct 1740
DB 1708 GATAGTCTAACTGGGATCTTATTAAGCATTAAATGATTTCTGGAGTAGCTTGACTGCT 1767
QY 1741 gttcctgcaaaactggcgacagcaactattcatattatctgtcccggttgcggaaaa 1800
DB 1768 GTTCTGCAAACTGGCGACACCACTTATTCATATTTATGCTGTCCCGTTGCCGAAAA 1827
QY 1801 gatgaccaagtattgattacttcatatgactaataatgaaatggagtagcgggttaaagga 1860
DB 1828 GATGACCAAGTATTAGTTACTTATATATGACTAATAGTAATGAGTAGCGGGTAAGGA 1887
QY 1861 atggattcaactgggaccgagttcttactacaataaaccgggataacaaactact 1920
DB 1888 ATGGGATCAACTTGGGACCGGAGTTCTTACTACAAATTAACCCGGATTAACCAACTACT 1947
QY 1921 gttttagctaaaatgactaaatcaaggagtaggttggattggagtagttcaagcaaaatct 1980
DB 1948 GTTTTAGCTAAATGACTAATCAAGGGATGGATTTGGGATGATTCAGCGGAAATCTT 2007
QY 1981 gatatgattggtttagatccgctcttaccctggcgaaacgtgataaaacctgttgat 2040
DB 2008 GATATGATTGTTGATTTAGACTCCGCTGCTTACCTGGCGAACGTGATAAACCTGTTGAT 2067
QY 2041 tgggacttaattggttggattaaacccgcatgactcgtctacacccaaatgatcctgaa 2100
DB 2068 TGGGACTTAATTGTTATGGATTAACACCGCATGATCTCTGCTACACCAAAATGATCCTGAA 2127
QY 2101 acgcaactacacagaaacccctgagacacctaatatctcccaaacaccaaagactcct 2160
DB 2128 ACGCAACTACACAGAAACCCCTGAGACACCTAATACTCCCAAAACCAAGACTCCT 2187
QY 2161 gaaatcctgggacacctcaactcctaatacacctatactcccggaattcctttaact 2220
DB 2188 GAAATCTGGGACACCTCAAACTCCTAATAACACCTAATACTCCGGAATTCCTTTAACT 2247
QY 2221 ccagaaacgcttaagcaacctgaaacccaaactaataatcgtttgcccacaaactgaaat 2280
DB 2248 CCAGAACGCTTAGCACTGAAACCCAACTAATAATCGTTTGCCCAAACTGGAAT 2307
QY 2281 aatgccataaagccatgattggcctagggtatgggaacattgcttagtattgttgc 2340
DB 2308 AATGCCATAAAGCCATGATTGGCCTAGGTATGGGAACATTGCTTAGTATGTTTGGTCTT 2367
QY 2341 gcagaaattaaacaacgtcgatttaac 2367
DB 2368 GCAGAAATTAACAACGTCGATTTAAC 2394

RESULT 3
LOCUS AX316616
DEFINITION Sequence 4 from Patent WO0190319.
ACCESSION AX316616
VERSION AX316616.1 GI:17899730
KEYWORDS Lactobacillus reuteri.
```

```
ORGANISM Lactobacillus reuteri
REFERENCE Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
AUTHORS Lactobacillus.
TITLE 1 (sites)
AUTHORS van Geel-Schutten,G.H., Rahaoui,H., Dijkhuizen,L. and van
Hiljum,S.A.
JOURNAL Novel fructosyltransferases
PATENT Patent: WO 0190319-A 4 29-NOV-2001;
FEATURES TNO (NL)
source Location/Qualifiers
1..2592
/organism="Lactobacillus reuteri"
/db_xref="taxon:1598"
1..67
/feature="Putative expression-regulating region"
RBS 29..32
RBS 54..57
misc_signal 2438..2592
/feature="Putative expression-regulating region"
BASE COUNT 896 a 465 c 501 g 730 t
ORIGIN
Query Match 100.0%; Score 2367; DB 6; Length 2592;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2367; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgtataaaagcgttaaaattggcgagtcgtttacactctcgactctcgctgctggtattt 60
DB 68 ATGTATAAAAGCGGTAAATAATGGCGAGCTGTTACACTCTCGACTCTCGCTGGGTATTT 127
QY 61 ggtgcacaaactgtaaatgcataccgcggacacaaatattgaaacaaatgattcttctact 120
DB 128 GGTGCACAACTGTAATAATGCATCCGGGACACAAATATTGAAACAATGATTCTTCTACT 187
QY 121 gtacaagtttacaacagtgataatgatattgtctgtaaaagtgtgacacttgtagtggt 180
DB 188 GTACAAGTTTACACACAGTGATATGATATGCTGTGTAAAGTGTGACACTTGGTAGTGGT 247
QY 181 caagtttagtcagctagtgatacgactatttagaactctctgtaactgtaaaagtctgct 240
DB 248 CAAGTTTAGTCAGCTAGTGATAGGACTATTAGAACTTCTGCTAATGCAATAGTCTTCT 307
QY 241 tctgcgcctaaatacacaaaattcttaacagtcagtcagtcagtcagtcagtcagtcagtc 300
DB 308 TCTGCCGCTAAATACACAAAATTTCTAAACGTCAAGTAGCAAGTTCTGCTGCAATAACATCA 367
QY 301 tctcaaaagttccgcagcttcattaaaataacacagatagtaaaagcgttaacgaatact 360
DB 368 TCTACAAGTTCCGCAGCTTCATTAAATACACAGATAGTAACGCGCTCAAGAAATAACT 427
QY 361 aatacagccaaatgatgacacgcgcaaaagcgtgcaccagcttaacgaatacttctgagct 420
DB 428 AATACAGCCAAAATGATGACACGCAAAAAGCTGCACCAGCTAACGAAATCTTCTGAAGCT 487
QY 421 aaaaatgaaccagctgtaaaacgttaagtattcttcagtcgcaaaaaatgatgatacaaa 480
DB 488 AAAAATGAACCAAGCTGTGAACCGTTAATGATTTCTTACGCTGCAAAAATGATGATCAACA 547
QY 481 tccagtaaaagaataactaccgcttaagtttaacaagagatgctgaaacgcttgtaaaaaag 540
DB 548 TCCAGTAAAAAGAAATATACCGCTTAAGTTAAACAGAGATGCTGAAACCGTTGTAAAAAAG 607
QY 541 gcgggaattgatccctaacagtttaactgatgaccagattaaagcattaaataagatgaac 600
DB 608 GCGGGAATTGATCCTTAACAGTTTAACTGATGACCAGATTAAAGCATTAATAAAGATGAAC 667
QY 601 tcttcgaagctgcgaagctggttacacaaaagactataatgattcccaaaagattgct 660
DB 668 TTCTCGAAGCTGCAAGGCTGTGTTACACAAATGACTTATATGATTTCCAAAGATTGCT 727
QY 661 gatacgttaatacaacaagatggttcggtacacagttccattctttaaagcaagtgaaatc 720
```



Db	728	GATACGTTAATCAAAACAAGATGGTCGGTACACAGTTCACATTCCTTTTAAAGCAAGTGAATC	781
QY	721	aaaaatagcttcgctcgctcacaaactaaagatgcacaaactaatactattgaacctttagat	780
Db	788	AAAAATACGCTTGGCTTACAACTTAAGATGCGACAAACTAATACTATTGAACCTTTAGAT	847
QY	781	gtatggatctcatggccagtttcaagatgcttcggacaggaacaagtctgctaattggaatgac	840
Db	848	GTATGGATTCATGCCAGTTTCAAGATGTTGGACAGGACAAGTCTAATTGGAAATGGC	907
QY	841	tatcaacttgcatcgcaatgatgggaattccaacaaataatgatcaatcatatctatctc	900
Db	908	TATCAACTTGTCTCGCAATGATGGGAATTCCAAACCAAATGATAATCATATCTATCTC	967
QY	901	ttatataatcaagtagtggtgaataagatgaagtcattggaagaatgaagttccaatttt	960
Db	968	TTATATAATAAGTATGGTGATAATGAATTAAGTCATTGGAAGAAATGATGGTCCAATTTT	1027
QY	961	ggctataattctacgggtttccaagaagatgctcaggatcgactgttttgaacagtcat	1020
Db	1028	GGCTATAAATTCTACGGCGGTTTCACAAAGATGCTCAGGATCAGCTGTTTGAACAGTGT	1087
QY	1021	aactctatccaatctattttatacaaggttagacacgtctgataacaataaccaatcaatcaa	1080
Db	1088	AACTCTATCCAAATTATTTATCAAGGGTAGACAGCTGTGATAACAATACCAAATCATCAA	1147
QY	1081	aaaatgctagcgtactctttttaaactgataataatggaaatgctatctaccaaacttatgataa	1140
Db	1148	AAAAATGTAGCGCTACTCTTTATTTAACTTGATAATAATGGAATGTATCACTGCTCTCAG	1207
QY	1141	gtacgaatgactatattgttgaaggtgatggctattactaccaaacttatgataa	1200
Db	1208	GTACGAATGACTATATTGTATTGAAGGTGATGGCTATTACTACCAACTTATGATCAA	1267
QY	1201	tggaaagctactaaacaaggtccgataataatgcaatgctgtagctccatgtaattgaa	1260
Db	1268	TGGAAGACTACTAAACAAGGTCCGATATATTTGCAATGCGTGATGCTCATGTAAATGAA	1327
QY	1261	gatgtaatggtagcgtaccttggttttgaagcaagtagctggtttggaaattatacaa	1320
Db	1328	GATGGTAATGGTGATCGGTACCTTGTTTTGAAGCAAGTACTGTGTTTGGAAAAATTATCAAG	1447
QY	1321	ggcaggaccacaatttataactggttcaaatattgaagctcgggcaacttgggctaaatgca	1440
Db	1388	GGCGAGGACCAAAATTTATAACTGGTTAAATATATGCGGAGATGACGATTTAATATCAAG	1447
QY	1381	agctatttagaattcttccaatgatgatattaagagctcggaacttgggctaatgca	1507
Db	1448	AGCTATTATTAGAAATCTTTCCAATGATGATATTAAAGACTCGGCAACTTGGCTAATGCA	1507
QY	1441	gctatcggtatctcacaactaaataaggaacaaaagaactctaagggtggcagagttatc	1500
Db	1508	GCTATCGTATCTCAAACTAATTAAGGACGAAAGAAATCTTAAGGTGGCAGGTTATAC	1567
QY	1501	tcaccattaatctgcaccaatggttaagcgatgaaattgagcgacacaaatgattgataa	1560
Db	1568	TCACATTAAATTCTCACCAATGCTAAGCGATGAATTTGAGCGACCAAAATGTAGTTAAA	1627
QY	1561	ttaggtaataataattactattttgcgctaccggtttaaatogaggaaagtaatgatgat	1620
Db	1628	TTAGTAATAATAATTACTTATTTCGCCGCTACCCGGTTTAAATCGAGGAACATAATGATGAT	1687
QY	1621	gcttggatgaatgctaaattatgcccgttggtagaattgtccaattgctcgagatgttgtct	1680
Db	1688	GCTTGGATGAATGCTAAATTATGCGCTTGGTGATTAATGTTCAATGGTTCGGATGTTGCT	1747
QY	1681	gatatctaacctggatcttataagccattaaatgattctggagtagctttgactgctctct	1740
Db	1748	GATAGTCTAACTGGATCTTATAAGCCATTAAATGATTCGAGTAGTCTTGACTGCTTCT	1807
QY	1741	gtccctgcaaacctggcgacagcaacttatcataattatgctgtcccccgttgcggaaaa	1800
Db	1808	GTTCCTGCAAACTGGCGACAGCAACTTATTTCATATTATGCTGTCCCGTTGCGGAAAA	1867

QY	1801	gatgaccaagctattagttacttcataatgactaataagaatagtagcggtataaagga	1860
Db	1868	GATGACCAAGCTATTAGTTACTTCATATATGACTAATAGAAATGGAGTAGCGGGTAAGGA	1927
QY	1861	atggattcaacttggccaccgagttbtttactactacaataaaccgcgataacacaaactact	1920
Db	1928	ATGGATTCAACTTTGGGACCGGAGTTTCTTACTACAAATTAACCCGGATACACAACACTACT	1987
QY	1921	gttttagctaaaaatgactaaacaaggagattggatttggatgatccaagcgaaaaatcctt	1980
Db	1988	GTTTTAGCTAAATGACTAATCAAGGGGATTGGATTGGGATGATTCAGACGAAAAATCTT	2047
QY	1981	gatatgattggtgatttagactcccgctgctttacctggcgaaactgataaaactgttgat	2040
Db	2048	GATATGATTGGTGATTAGACTCCGCTGCTTTACCTGGCGAAGCTGATTAACCTGTGGAT	2107
QY	2041	tgggacttaatttggttattgattataaaacgcgatgctctgtacaccaaataatgctcgtgaa	2100
Db	2108	TGGGACTTAATTGGTTATGGATTAAACCCGCATGATCTCTATACACCTAATCTCCGAAATTCCTTAACT	2287
QY	2101	acgccaactacaccgaaacccctgagacaccttaactcccaaacacacaagaactcct	2160
Db	2168	ACGCCAACTACACCGAAACCCCTGGAGACACCTAATACTCCCAAAACACCAAGACTCCT	2227
QY	2161	gaaatcctgggacacctcaaaactcctcaatacaccttaactcccggaattcctttaact	2220
Db	2228	GAAATTCCTGGGACACCTCAAACTCCTATACACCTAATACTCTCCGAAATTCCTTAACT	2287
QY	2221	ccagaaacgcttaagcaacctgaaacccaaactaataatcggtttccacaaactggaat	2280
Db	2288	CCAGAAACGCCTAAGCAACCTGAACCCCAACTAATATCGTTTGCCACAACTCGGAAT	2347
QY	2281	aatgccataaagccatgattggcctaggtatgggaacattgcttagtatgtttggtctt	2340
Db	2348	AATGCCAATAAGCCATGATTGGCCTAGTAGTATGGGAACATTGCTTAGTATGTTGGTCTT	2407
QY	2341	gcagaaattacaacacgctcgatttaac	2367
Db	2408	GCAGAAATTACAACGTCGATTTAAC	2434
RESULT	4		
STRFTF	STRFTF	4305 bp	DNA
LOCUS	S.mutans fructosyltransferase gene, complete cds.		linear
DEFINITION	M18954		BCT 15-FEB-1996
ACCESSION	M18954		
VERSION	M18954.1	GI:153635	
KEYWORDS	fructosyltransferase.		
SOURCE	Streptococcus mutans (strain GS5) (clone: pSS22 and pTS102.) DNA.		
ORGANISM	Streptococcus mutans		
REFERENCE	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;		
AUTHORS	Shiroza,T. and Kuramitsu,H.K.		
TITLE	Sequence analysis of the Streptococcus mutans fructosyltransferase gene and flanking regions		
JOURNAL	J. Bacteriol. 170 (2), 810-816 (1988)		
MEDLINE	88115184		
COMMENT	Draft entry and printed copy of sequence for [1] kindly provided by H. K. Kuramitsu, 06-MAR-1988.		
FEATURES	Location/Qualifiers		
source	1. .4305		
	/organism="Streptococcus mutans"		
	/strain="GS5"		
	/db_xref="taxon:1309"		
	/clone="pSS22 and pTS102."		
	26. .394		
	/note="ORF1; putative"		
	/codon_start=1		
	/transl_table=11		
	/product="unknown protein"		
	/protein_id="AA88583.1"		
CDS			



```

/db_xref="GI:1196925"
/translation="MSRILEELKVVIRPEFINRIDEKVVFNHNGQEDIRHVVKIMYA
PLAHLAADQGITIKFPQSLAKHIALAGYDAEMGARPRLRTLQTEVEDKLAELTSLGSL
ASQGTUKIGIQKEKLAFDIV"
681..782
sig_peptide
CDS
681..3074
/note="precursor"
/codon_start=1
/transl_table=11
/product="fructosyltransferase"
/protein_id="AAA88584.1"
/db_xref="GI:153636"
/translation="MTETKVRKMYKKGKFWVATITAMLTIGLSSVQADSEANSTQV
SSELARSSQVQENTTASSNAENQAKTEVQETPTNPAATVNTDOTTITIDNAAV
ESKASTKQDAATVTTAASTPEVGQTNKEDKAKATEADITTPKNNTIDBYGITEQAR
KITEAGINLSSITQKVALNKVLKLTSDAQTHQMTYQFEDKIAQTLIAQDRYAIAT
YFNAKAIKNMKAAITRDAQTQADLQVWDSWPQDAKTGEVINWNGYQLVAMMGTP
NTDKNHYYLLNRYGDNNTDHMKNAGSIFCYNETPLTQEWSGSATVNEGSLQFLYTK
VTDSDKSNRORLATATVNLGPDQDQVILLSVENDKVLTPEGVMAYHYOSYQWRSTF
TGADNADMYRPHVIEDNGRVLVFEASTGTENYQGEDQIYNFTNTGGSSAYNVKSLF
RFLDDQDMYRNASWANAIGILKGDKKTPEVDQFYTELSSSTMVSDLELRPNVKVL
GDKYIFTLASRLNHSNDAWNKANENVGDNVVMGVSDQLTNGYKPLNNSGVLTIA
SPVADNRTATSYIAYFVAGSSDSTLLMTAYMTNRRNEVAGKKNSTWAPSLIQLVPD
TTVKLAEMTQQQGDWIDEPSRTDITVTGLDTAYLPGENGYIDWNVIGGYGLKPHPTP
QYQTPVPSYPIHPDDIIISPEVSPDGHVLPVKVNNDSAGRIDQSRNSGSGSLNAVNV
SAGGNI SVKPSQKS INNTRETKKAHVSTEEKQKKGNSGFFAALLALFSAFCVSIGK"
783..3071
mat_peptide
terminator
CDS
3150..3179
/product="fructosyltransferase"
/note="putative"
/complement(3265..3951)
/notes="Orf3; putative"
/codon_start=1
/transl_table=11
/product="unknown protein"
/protein_id="AAA88585.1"
/db_xref="GI:1196926"
/translation="MFSQKLKERRKKLGLSOAQTADKLGISRPSYFNWEIGTKTKPN
KNLDKLAHLKVDASVFLSHQDITVEITYRLNESNKTLYKYSQHLLPEODDKKRLMKN
KRYPVREKLKSAGTCYSYSGDGNFTVEYDEIDHDFASWIFGDSMEPIFLNGEVAL
IKQGTGDEGAIYAIWDQGTQYIKKKYREETGLRLVSLNKKADKPAFDYDENPRIIGL
IVGNFIPLLEG"
CDS
4269..54305
/notes="Orf2; putative"
/codon_start=1
/transl_table=11
/product="unknown protein"
/protein_id="AAA88586.1"
/db_xref="GI:1196927"
/translation="MANKKIRLRKA"
BASE COUNT 1420 a 702 c 819 g 1364 t
ORIGIN 1 bp upstream of HindIII site.

```

Query Match	30.4%	Score 718.6	DB 1	Length 4305
Best Local Similarity	62.2%	Pred. No. 4.5e-124		
Matches 1167	Conservative	0	Mismatches 699	Indels 9
Gaps	2			
Qy 247	gctaatacacaaaattctaacagctcaagtgacgcaagttctgctgcaataacatcatctaca	306		
Db 885	GCTAAGACTGAAGTTCAGAAACTCCATCAACAATCCCGACCTGCTACTGTTGAGAAC	944		
Qy 307	agttccgcagcttcttaataataacacagatagtaaaagcggctcaagaataataataata	366		
Db 945	ACTGATCAACAACAATAAGGTGATAACAGATAAATGCTGCTGTAATCAAAAGCAAGTAA	1004		
Qy 367	gccaaaaatgatgacacgcgcaaaaagctgcaccagctaacgaaattcttgaagctaaaaat	426		
Db 1005	ACTAAGGATCAACGAGTACCGTAACTAAACAGACGCTAGTACACCGGAAG---	1061		
Qy 427	gaaccagctgtaaacgttaatgattcttcagctgcaaaaaatgatgatcaacaatccagt	486		
Db 1062	CAACAATAATCAAAAGGATAAGGTTAGGCTAAGGCAACTTAAGAAGCTGACATAAATACGCCAAAG	1121		

Qy	487	aaaaagaataactaccgcgtcaagtttaacaagaagatgctgtaaaaaacyttgtataaaaaagcgcgga	546
Db	1122	AATACAATAGATGAATACGGCTTAACAGAACACAGGCTCGTAAGATTCGCTACTGAAGCTGGT	1181
Qy	547	attgatcctaacagtttaactgatgccagattaaagcattaaataagatgaactcttcg	606
Db	1182	ATPAAATTTAGAGAGTTTGACACAAAACGAAGTTGAAGCAATTAATTAAGTTAAATTAACG	1241
Qy	607	aaagctgcaaaagtctggtcacacaaatgactataatgatttcccaaaagattgctgatcag	666
Db	1242	AGTGATGCTCAAAACGGGTCACTAAATGACCTATCAAGAATTTTGACAAGATTGCTCAACAG	1301
Qy	667	ttaatcaaaaagaatggtcgggtacacagttccattctttaaagcaagtgaaatcaaaaat	726
Db	1302	TTGNATGCTCAAGTGAACGCTATGCTATGCCCTATTTTAAATGCAAAACGAATCAAAAAT	1361
Qy	727	atgctgcgcgtacaactaaagatgcacaaaactaatactattgaaacctttagatgtatgg	786
Db	1362	ATGAAGCGGCTACAAACGGGTGATGCCAAACGGGTCAAATAGCTGATTGGGATGTTTGG	1421
Qy	787	gattcatggccagttcaagatgttcgacagcagagaagttgctaattggaatggcctatcaa	846
Db	1422	GATTCCTTGGCCAGTTCAGGATGCTGAAGCTGCTGAAGTTTAAATTTGAAGTGGTATTCAG	1481
Qy	847	cttgtcatcgcaatgatgggaattccaaacccaaaatgataactatcatctctctctat	906
Db	1482	CTTGTCTGCTATGATGGGCATCCAAATACTAATGATAAATCATATTTATCTCTCTTTAT	1541
Qy	907	ataagtatgggtgaataatgaattgaatgctattggaagaatgtaggtccaattttggctat	966
Db	1542	AATAAATATGGAGATAATAATTTTGATCATTTATGGAAAAATGCAAGTCTCTATCTTGGTTAT	1601
Qy	967	aattctaccgcggttcacaagaatggtcagatcagctgttttgaacagtgataactct	1026
Db	1602	ANTGAACACCCCTAACTCAAGAATGGTCAGGTTACGCTACCGTAATCAAGATGGGAAGT	1661
Qy	1027	atccaattatttatcaaggggtagacacgctctgataaccaatcaccaatcatcaaaaaatt	1086
Db	1662	TTGCAGTTATTCTACACCAAGGTTGATACTAGTCACAAAAACAGTAACAATCAACGTTTA	1721
Qy	1087	gctagcgtactcttatttaactgataataatggaatgtatcaactgcctcaggtaaga	1146
Db	1722	GCAACAGCGACTGTAATCTTGCGTTTGATGACCAAGATGTTAGAAATTTCTTCTGTGAA	1781
Qy	1147	aatgactatattgta-----tttgaaggtgtagtgcctattactaccaaaacttatgatcaa	1200
Db	1782	AATGATAAAGTTTTTAACGCTGAAGCGGTGATGGCTATCATTTATCAAAGTTATCAACAA	1841
Qy	1201	tggaaagcctaatacaaaaggtccgataatatgcaatgcgtgtagtgcctatgtaattgaa	1260
Db	1842	TGGCGTTCAACCTTTACAGGTGCTGATAATATGTATGCTGCTGATCCACATGTCTATTGAA	1901
Qy	1261	gatggtaatggtgatcgtaccctgtttttgaagcaagtaactggttttgaaaaattatcaa	1320
Db	1902	GATGAGAAATGAGATCGCTATCTTGCTTTGAGGCTAGTACAGGTCACAGAAATTTATCAA	1961
Qy	1321	ggcgaagaccaaaattataactggtttaaatatgctggcgagatgacgcatttaatatcaag	1380
Db	1962	GGTGAAGATCAGATTTTACAACCTTTACTAACTATGGCGCAGCTCTGCTTTATAATGTTAAA	2021
Qy	1381	agctattttagaattctttcccaatgatgatatgaagtcggtcgggaacttgggtcaatgca	1440
Db	2022	AGTCTTTTATGATTTTATGATGATCAAGATATGTTATACCGTGGCAAGCTGGGCCAATGCA	2081
Qy	1441	gctatcggttatctctcaactaaataagggacgcaaaaagaatcctaaggtggcgaggttatc	1500
Db	2082	GCTATTGGTATTTTAAACCTTAAGCGCGATATAAAAACACCTGAGGTAGATCAATTTTAC	2141
Qy	1501	tcaccatattaattctgcaccaaattgtaagcgatgaaattgagcgaccaaaatgtagttaaa	1560
Db	2142	ACGCCCTTTACTAAGTTTCAACAATGGTTTCGGATGAACCTCGAGCCACCAATGTGTTAAA	2201

QY 1561 ttaggtaataataattacttatttgcgcgtaccgcttttaaatcaggaagaaatgatgat 1620  
 Db 2202 TTAGGAGATTAATATCTTTTACAGCTTACGCTTTAAATCACGGAAGTAACACGAT 2261  
 QY 1621 gcttgatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 1680  
 Db 2262 GCTTGGAAATGAACAAATGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2321  
 QY 1681 gatagcttaactgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 1740  
 Db 2322 GATCAATATGACTAACCGCTTACAAACCTTAAATATAGTGGTGTGTGTGTGTGTGT 2381  
 QY 1741 gttcctgcaactgagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1800  
 Db 2382 GTTCCAGCAGATGGCGAAGCGGACTTACTCTTATATGCTGTGTGTGTGTGTGTGTGT 2441  
 QY 1801 gatgacaaagattagttacttactatgactaataagaaatgagtgagtgagtgagtgag 1860  
 Db 2442 TCTGATACCTGTATATGACGCTTATATGACCAATCGTAATGAAGTGCAGGTAAGGT 2501  
 QY 1861 atggattcaactgggacccgagtttacttactacaaataaccccgagataacacaaactact 1920  
 Db 2502 AAGATTCACCTGGGCACTAGTTTCTGATTCAGGTTTACCAGATGGGACTACAAA 2561  
 QY 1921 gttttagctaaaatgactaatacaagggatggattggatggatggatggatggatggatggat 1980  
 Db 2562 GTCTTAGCAGAAATGACACACAGGGTGTATGGATTTGGGATGAACCAAGTCGCACGACA 2621  
 QY 1981 gatagtgattgattagactccgctgtcttaccctgagcagcagcagcagcagcagcagcagcag 2040  
 Db 2622 GACACTGTGGCAGCTGATACAGCTTATCTCCAGGTGAAATGATGTTATATTCAT 2681  
 QY 2041 tgggacttaattgattgattgattgattgattgattgattgattgattgattgattgattgatt 2100  
 Db 2682 TGGATGTTATGGTGTAGCGTTTGAAGCCACATACACCGGGACAATATCAACCAACT 2741  
 QY 2101 acgccaactacacca 2115  
 Db 2742 GTTCCATCAACACCA 2756

## RESULT 5

AX316622  
 LOCUS AX316622 4634 bp DNA linear PAT 14-DEC-2001  
 DEFINITION Sequence 10 from Patent WO0190319.  
 ACCESSION AX316622  
 VERSION AX316622.1 GI:17899731  
 KEYWORDS Lactobacillus reuteri.  
 SOURCE Lactobacillus reuteri.  
 ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 1 (sites)  
 van Geel-Schutten, G.H., Rahaoui, H., Dijkhuizen, L. and van  
 Hijum, S.A.  
 TITLE Novel fructosyltransferases  
 JOURNAL Patent: WO 0190319-A 10 29-NOV-2001;  
 TNO (NL)  
 FEATURES  
 Source 1..4634  
 /organism="Lactobacillus reuteri"  
 /db\_xref="taxon:1598"  
 1205..1210  
 1220..3598  
 /note="unnamed protein product"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAD19335.1"  
 /db\_xref="GI:17899732"  
 /translation="MYKVGKNWAVATLVASILMGVVTVAHADVFNNGVAEYNT  
 EROANGOIGVDGRIISANTSGSTNOESSATNTENAVVNESKNTNTNENAVNEN  
 KNTNTNENAVNENKNTNTENDNSOLKITNEOPSAAOANLKNLPQAAQVNAK  
 IDAGSLTDDOINELNKFINSKSAEKGAKLTFKDLGIGNAIVKQDPQYAIPIYSNAKEI

KNMPATYTVDAOTGKWAHLVDWSDHPVODPYTVSYNSYNGVQLVYVIMGIMPNSPTGDN  
 HYLXLYNKYGNDFSHRNAGSIFGKTNVQFQWSSGSAIVNDGDTIOLFSTNSDTSD  
 YKLDRLATATLNLVNDGVSISVDNYOVLFEFGDFHYOTYBQFANGKDRNDYD  
 CLRDPHVLENGDRYLVFEANTGTDYSDQIYNWANYGDDAFNFKSFKLLNNK  
 KRELASGLANGALGILKLTNNQSKPVEEYSPVSTLMGDEVKXKGLQDYLLFSVT  
 RVRSGLRELKATONTIVGDNVAMIGYVSDSLMGKYKPLNNSGVVLTASVPANMRTAT  
 XSYAVPVAGHPDOVLITSYMSKDFASGEGNYATWAPSLVQINPDYTVILARATN  
 OGDVMDSSSRNDNMLGVLEGAASALPGEWKPVDWLSLNSKSPGTLPHQPVQP  
 KIDQDQOPSGONTKNVTPGNDPAGKATPDNTNIDPSAQFNTIDPSAOSXGO  
 NTKNTPGNEKOGNTDAKQLPQTNKSGLAGLSLALFLAALAEKRAH

BASE COUNT 1530 a 837 c 897 g 1345 t  
 ORIGIN

Query Match 26.9%; Score 637; DB 6; Length 4634;  
 Best Local Similarity 59.0%; Pred. No. 6.7e-109;  
 Matches 1266; Conservative 0; Mismatches 821; Indels 60; Gaps 8;

QY 114 ttctactgtacaagttacaacagggtgataatgatatgtctgtttaaagtgacactgg 173  
 Db 1273 TTTAATGGGAGGGGTGTAAACCGCTCATGCTGATCAAGTAGAAGTAACATTAACAAG 1332  
 QY 174 tagtggtcaagttagtcagctagtcagctagtcagctagtcagctagtcagctagtcag 233  
 Db 1333 TGTTGCTGAAGTTAATCTACTGAACGCTAAAGCTAAAGCTAAAGCTAAAGCTAAAGCT 1392  
 QY 234 tgcttctctgcgcgttaataacacaaattctaacagtcagtcagtcagtcagtcagtcag 293  
 Db 1393 TATTAAGTCTTAACAGTAAATACACAGTGGCTGCGAAGTAACAGTAATCATCTGTACTAA 1452  
 QY 294 aacatcatctcaagttccgcagcttccatataacacacagatgataagcggctccaaga 353  
 Db 1453 CA---ATACTGAAAAATGCTGTGTTAATGAAAGCAAAATACTACATACTGAAATGTC 1509  
 QY 354 aataactaatcacgcaaaaatgatgcacgcacacacacacacacacacacacacacacac 413  
 Db 1510 TGTTGTTAATGAAACAAAAATACTAAACAAATCTGAAAAATGCTGTTTAATGAAACAA 1569  
 QY 414 tgaagctaaaaatgaaccagctgttaacgcttaagcttccagctcagctcagctcagctcag 473  
 Db 1570 AAATACTAACACACAGAAACGATATAGTCAATTAAGTTAACT---ATAATGAACA 1626  
 QY 474 tcaacaatccagttaaaagaatactaccgcgttaagtaaacacagatgataagcggcttgc 533  
 Db 1627 ACCATCAGCCGCTACTCAAGCAAACTTGAAGAAGCTAAATCTCAAGCTGCTAAGGCTGT 1686  
 QY 534 aaaaagcggggaatgtatccttaacagcttcaacagtcacacacacacacacacacacacac 593  
 Db 1687 TCAAAAATGCCAAGATGTGCGGCTAGTTTACAGATGATCAAAATTAATGAATTAATAATA 1746  
 QY 594 gatgaactctcgaagctgcaaaagctggttacacaaatgacttataatgatttccaaaa 653  
 Db 1747 GATTAACCTCTCTAAGTCTGCTGAAAGGGGTGCAAAATTTGACCTTAAAGGCTTAGAGGG 1806  
 QY 654 gattgctgatacgttaatacaagaagatggtcggtacacagcttccatttcttaagaagaag 713  
 Db 1807 GATTGGTAATGCTATTGTTAAGCAAGATGCCAATATGCTATTCTTATTCTTATTCTTAA 1866  
 QY 714 tgaatcaaaaaatgctgcgcgtacacaaatgataagatgacacaaactaactattgaacc 773  
 Db 1867 GGAATCAAGAATATGCTGCAACATACACTGTAGATGCCCAACACAGTAGATAGGCTCA 1926  
 QY 774 tttagatgattgattcattcagccagtcgaagtgttcgcagacaggaagttgctaattg 833  
 Db 1927 TCTTGATGCTGGGACTCTTGGCCAGTACAAGATCCCTGTCACAGGTTATGATCTAATA 1986  
 QY 834 gaatggcctatcaactgtcgcgtcgaatgaggaattccaaaa-----caaaatgataa 887  
 Db 1987 CATGGGTATCAACTAGTATTGCTATGATGGGTATTTCAAAATTCGCAACTGGAGATA 2046  
 QY 888 tcatatctatctcttataataagtaagttgataagtaagtaagtaagtaagtaagtaagtaag 947  
 Db 2047 TCATATCTATCTCTTTTACAAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2106

Qy	948	agg	tccaatttttggctataaattctaccgcggtttccacaagaatggtcagatcagctgt	1007
Db	2107	AGG	TTCAATCTTTTGGAACTAAAGAAACAAATGTGTTCCAAGANTGCTCAGGTTCACTAT	2166
Qy	1008	ttg	aacagtgataaactctctatccaattattttatacaagggtagacacgtctgataacaa	1067
Db	2167	TGT	AAATGATGATGGTACAAATTAACAATAATTTTCACTCAAAATGATAGTCTGATTAACA	2226
Qy	1068	tacca	atcatcaaaaaatttgttagcgtactcttatttaactgetaataatgaaaatgt	1127
Db	2227	GTT	GAAATGATCAACGCCTTGCTACCGCAACATTAACCTTAATGTTGATGAACGGTGT	2286
Qy	1128	atac	tcgctcaagtaacgaataagactatatgtattgaagtgatggcttactaccac	1187
Db	2287	TTCA	ATCAGAGTGGTTGATTAATATACAAGTTTGTGTTTGAAGTGATGGATTCCTACTACCA	2346
Qy	1188	aact	tatgatcaatggaagagctactaacaagaagtgc---ccataaatatgcgaatgcgtga	1244
Db	2347	AAC	TTATGAACAATTCGCAACGCAAGATCGTGAAATGATGATTACTGCTTACGTGA	2406
Qy	1245	tgct	oaagtaagatggaatggtgatcggttacctgttttgaagcaagactcgg	1304
Db	2407	CCCA	CACGTTGTTCAATATAGAATAATGGTGATCGTATCTTGATTCGAAGCTAATACTGG	2466
Qy	1305	tttg	aaaaattatacaagcgagaccaaattataaactgggttaattatgcgcgagatga	1364
Db	2467	GAC	AGAAGATTACCAAAAGTCGACACCAATTTATATTTGGCTAACATGTTGGTGGCGATGA	2526
Qy	1365	cgatt	taatacaagagacttattagaattcttccaatgatgatataagagtcgggc	1424
Db	2527	TGC	TTTCAATATATAAGAGTTTCCTTCAAGCTTTTGAAATAAAGAAGGATCGTAATTGGC	2586
Qy	1425	aact	tgggcgaatcagctatcagttactcctcaactaaataaagagcgaagaatcctaa	1484
Db	2587	TGG	TTTAGCTAATGGTGACCTGGTATCTTAAAGCTCACTAACAATCAAAAGTAAGCAAA	2646
Qy	1485	ggtg	cgaagttatactcaccattaaattctgcaccaatggaagcgatgaattgagcg	1544
Db	2647	GGT	TGAAGAAGTATACTCACCATTTGGTATCTACTTTCATGGCTTCGATGAGCTA----	2701
Qy	1545	accaa	atgagttagtaataatagtaataattacttattgcgcgtaccogtttaaatcg	1604
Db	2702	-----	NNNNNAAGCTTGGTGATAGTATATCTCTCCGTAACCTGGTGAAGTCG	2754
Qy	1605	agga	agtaatgatgacttggaatgactaaattatgcogtttggtgataattgttgcgaat	1664
Db	2755	TGT	TTCCGATCGTGAATTAACCGCTGAAGGATAACACAATCGTTGGTGATACGTTGCTAT	2814
Qy	1665	ggtc	ggatgatgttgcgatagtctaaactgatcttataagccaattaaatgattctggagt	1724
Db	2815	GAT	TGGTTACGTTCCGATAGCTTAATGGGTAAGTACAAGCCATTAAATAACTCAGGTT	2874
Qy	1725	agtc	ttgactctctctgcgcgaactggcgagacgaacttattcattatattgctgt	1784
Db	2875	CGT	ATAAATCGATCAGTACCTGCGAAACTGGCGTACTGCTACTTATCTACTATGCAAGT	2934
Qy	1785	cccc	gttcggaaaaagatgacaagtattagttacttcatatatgactaatagaaatgg	1844
Db	2935	ACCT	GACTGGTGCATCTGATCAAGTATTAATTAATTTACTTCTTACATGAGTAACAAGACTT	2994
Qy	1845	agta	cggtgtaaaaggaattgattcaacttgggcacgcgagttcttactacaattaaacc	1904
Db	2995	TGCT	TACGTTCAAGGAAACTATGCAACTTGGGACCAAGTTTCTTAGTACAATAACATCC	3054
Qy	1905	gga	taacacaaactactgttttagctaaaaatgactaatcaaggggattggattgggagta	1964
Db	3055	AGAT	GACACGACAACACTGATATAGCAGTGCACATAACCAAGGTACTGGGTGTGGACGA	3114
Qy	1965	ttca	gcgaataactctgatattgagtgattt-----agactccgctgc	2009
Db	3115	CTCT	AGTCGGGAACGATTAATATGTCGGTGTCTTAAAGAAGGTGCAGTAACAGTGC	3174

QY	2010	tttacctggcgaaacgtgataaaacctctgattgggacctaaat-----tggtta	2057
Db	3175	CTTACCAGGTGAATGGGTGAAGCAGCTTGACTGGAGTTTGATTAAACAGAAGTCTGGCTT	3234
QY	2058	tggtataaaacgcgatctcgtctacacaaatgatcctgaaacgccaactacaccaga	2117
Db	3235	AGGCTTAAGCCCTCATCAACCACTTCAACCAAAAGATTGAT-----CAACCTGATCAACA	3288
QY	2118	aacccttgagacacacctaataactccccaaacacacaaagactcctgaaaatcctctgggacacc	2177
Db	3289	ACCTTCTGTGTCAAAACACTAAGAAATGTCACACCAAGTAATGGTGATAGCCGTGCTGGTAA	3348
QY	2178	tcaaaactcctaatacacttaatactcctcggaatactcttaactccag	2224
Db	3349	GGCAACTCTGATCACTAAATATTGATCCAAAGTCACAAACCTTCTG	3395
RESULT	6		
LOCUS	STRTFPB	3600 bp	DNA linear
DEFINITION	Streptococcus salivarius fructosyltransferase (ftf) gene, complete cds; uracil phosphoribosyltransferase gene, 5' end; IS1161 transposase gene, 3' end.		
ACCESSION	L08445		
VERSION			
KEYWORDS			
SOURCE	L08445.1	GI:348000	
ORGANISM	Streptococcus salivarius.		
REFERENCE	Streptococcus salivarius.		
AUTHORS	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae.		
TITLE	Streptococcus.		
JOURNAL	1 (bases 1 to 3600)		
MEDLINE	Rathsam,C., Giffard,P.M. and Jacques,N.A.		
REFERENCE	The cell-bound fructosyltransferase of Streptococcus salivarius: the carboxyl terminus specifies attachment in a Streptococcus gordonii model system		
AUTHORS	J. Bacteriol. 175 (14), 4520-4527 (1993)		
JOURNAL	9322332		
MEDLINE	2 (bases 1 to 3600)		
REFERENCE	Jacques,N.A.		
AUTHORS	Direct Submission		
TITLE	Submitted (04-JAN-1993) Institute of Dental Research, Australia		
JOURNAL	Location/Qualifiers		
FEATURES	1..3600		
SOURCE	/organism="Streptococcus salivarius"		
CDS	/db_xref="taxon:1304"		
	/tissue_lib="lambda-A39/ATCC 25975"		
	<1..56		
	/standard_name="IS1161"		
	/function="transposition"		
	/note="ORF4; putative"		
	/codon_start=3		
	/transl_table=11		
	/product="transposase"		
	/protein_id="AAA71924.1"		
	/db_xref="GI:348001"		
	/translation="IHGYQSAKKLFELTQTA"		
gene	220..3129		
	/gene="ftf"		
CDS	220..3129		
	/gene="ftf"		
	/function="sucrose cleavage and fructan synthesis"		
	/codon_start=1		
	/transl_table=11		
	/evidence="experimental		
	/product="fructosyltransferase"		
	/protein_id="AAA71925.1"		
	/db_xref="GI:348002"		
	/translation="MDITVNSOSNTVAPKQAECKMRYISIRKVAIVTGATSLVGTALF		
	LGATOVKQDVQVETAPAVATATPETSTASLTAVSETATSVATSVASSVVAHSEVA		
	TKPVETQPSNTSPSVVERKASTVTSDDATTPSATVAASPAHTSEAAVEAPTS		
	TASSEADHTEVDLVKVSNSAANSLKNGRIKSIVENMTSDQIVALTBEETKAL		
	NKVRSDDAIKGTGTSLVTRNLKDIVASFLLKQDSKLAVPYFKADTIINNPATVDAQ		
	TWKBEIDVWDSMPVDQASGVVSNWNGQLVISMAGAPNKNNSHLYLLRYKYGNDIE		

THWKNAGLEFGVNALEDDDOORSGSATVNSDQSIQYLYTKNDTSGGKLNWQOLASATLN  
 LAIVENDEVKSVENDHILFGDNGHYOSYKFMSTFDDHNDHCNPDRONTYCLRDP  
 HLIENGSRILYFESNTDQNGEYKMSYNGDDDAFNKLSFNTVNNKHLNLA  
 SWANGSIGILLDDNEKPSVAELVPLVTSMTVDEVERPSVVKMGSKYLYFTASRI  
 NKSTDAEGTVAAREAVGDVVMGLFVSDLSKRYPLNGSVGLMGTVPADPRTSTYS  
 YVAVPEGSDDTLVTSYMTNRGGIAGAKENSTWAPSLIKHNADDTTEVLPMKTNQG  
 DWIMDKSESILVHVGSDNSAKLPNEDFNVDYAVSGYGLKPHITPTVDGSGVSEAHG  
 VLTVTVDGKUKADKAPETPVSPTEGNHSDVDKTKNPKTSKPDNNDQPSADKEDKPTN  
 PTPDPSARTFPPYIGDHSNDNSNDHVAVPVKPSTGDSVGDORRVPVQAAREIATPV  
 KNTIVAGTPTVPTNTVKEESVTEAPKPVSEKQVSHGVKANEVTKSDSESKGNN  
 TKVAALKLATTPKTPSDSEGSNLSILATIFAASLALIGYGLVTKIHLPKK"  
 3219...>3600

CDS

/EC\_number="2.4.2.9"  
 /function="formation of uridine 5'-monophosphate from  
 uracil"  
 /note="ORF2: putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="uracil phosphoribosyltransferase"  
 /protein\_id="AA071926.2"  
 /db\_xref="GI:12597937"  
 /translation="MGKQVLSHPLIQHLKLSILREDTSTKDFRELVNEIAMLGYEV  
 YRDLPEVEIOTITTKTVOKQLSKLAIVIPILRAGIMVDFLSLPAKRVGHIGM  
 YRDEETLEPVEYLVKLPEDIDQRI"  
 BASE COUNT 1178 a 719 c 734 g 969 t  
 ORIGIN

Query Match 23.2%; Score 548; DB 1; Length 3600;  
 Best Local Similarity 58.2%; Pred. No. 2.5e-92;  
 Matches 1021; Conservative 0; Mismatches 700; Indels 33; Gaps 2;

Qy 376 gatgacgcgcaaaagctgcaccagctaaacgaatctctgaagctaaataaagcagct 435  
 Db 661 GCTCATACATCAGAGCTGCGAGTAGAACCCAACTTCAACAGCAAGCTCAGAGCAGCT 720  
 Qy 436 gtaacgcttaagtattctcagctgcgcaaaatgatcatcaaatccagtaaaagaat 495  
 Db 721 GATACACACACTGAAGTAGACCTTAAAGTTCTGAAAACACGCGGCTTAATGCCAACCCT 780  
 Qy 496 actacgcgtgaagttaaacagagatgcgaaacgttgtaaaaaagcgagggaattgatcc 555  
 Db 781 AGTAACTTAATGGTCGTATCAAGTCTATTGCTTGAAGAGAATATGACCTCAGATCAAAATC 840  
 Qy 556 aacagtttaactgatgacagattaaagcattaaatgaagtgaaactctcgaagctgca 615  
 Db 841 GTTGCTTACTGAGGAGAAATTAAGCACTTAATAAGTTGACTTTAGTGATGATGCT 900  
 Qy 616 aagctcgtgacacaaa-----tgactataatgattccaaagattgctgatacgtta 669  
 Db 901 ATCAAGGGTACAGGTACTAGCTTGACTTACCGTAACCTTGAAGATATATCGTTGCTAGCTTC 960  
 Qy 670 atcaacaagaatgctcgtgacacagctccattctttaaagcaagtgaaatcaaaaatag 729  
 Db 961 TTGAAGCAAGACACAAATTTGGCAGTTCCYTTACTTTAAACGACATCAATCACTAATG 1020  
 Qy 730 cctgcgcctcaacataaagatgcacaaactaactattgaaaccttttagatgattggat 789  
 Db 1021 CCTGCCCTTCAACACTGTGTGATGTCGCAACAGCATGAAAAGAAAGAAATCGACGTTTGGGAT 1080  
 Qy 790 tcatggccaggttcaagatgctcggacagcacaagttgctgaattggaatggctatcaact 849  
 Db 1081 TCTTGGCCACTCAAGATGCTAGTCAAGGTAGTTAGTACCACTGGAATGGTTACCAAGTTG 1140  
 Qy 850 gtcacgaatgatgggaatttccaaaccaaataatgataatcatatctatctctctataat 909  
 Db 1141 GTTATCTCAATGGCGGTGACCAAAATAAAATTAACACCAATTTACTTGTCTCTATAGA 1200  
 Qy 910 aagatggtgataatgaattaaatcattggaagaatgtaggtcccaatttttggctataat 969  
 Db 1201 AAATATGGGGAATATGACTTTTACTCATTTGGAAAAATTCGGGGTCCCAATCTTTGGTTACAAT 1260  
 Qy 970 tctacgcggttttcacaaagaatggctcaggatcagctggttttgaaacagtgataactctac 1029

Db 1261 GCCCTTGAAGATGACCAACAAATGGTCAGGTTCCAGGACTGTTAACTCTGATGGTAGTATC 1320  
 Qy 1030 caattattttacaaagggtagacacgctctgtataaacaataaccaatcatacaaaaaattgct 1089  
 Db 1321 CAACCTTTATACACTAAGAAATGATACTAGTGGTGTAAATTAACCTGGCAACAATTGGCT 1380  
 Qy 1090 aqcgctactctttatttaactgataataatgaaatgatacactcgcctcagtcagtaacaaat 1149  
 Db 1381 AGTGCTACACTTAACTTGGCTGTGAAAACGACGAAGTTGTTATCAAGTCAGCTTGAAAAT 1440  
 Qy 1150 gactatttatttgaaggtagctgtattactaccacaaacttatgatacaatggaagact 1209  
 Db 1441 GACCACATCTCTTTGGTGGAGATAATATCACTATCAAGACTATCTTAATTTATGAGT 1500  
 Qy 1210 actaacaagggtgcgagata-----atattgcaatgcgt 1242  
 Db 1501 ACTTTTGTATGATGACCATTAACCATGATGGTAATCCAGACCGTACGGGATAACTACTGTCTT 1560  
 Qy 1243 gatgctcatgtaattgaagatggttaattgtagatcggtacactgtttttgaagcaagtaact 1302  
 Db 1561 CGTGACCCACACATCATTTGAAGATAACGCTAGCCGTTACCTTATTTTGAATCAAAATACA 1620  
 Qy 1303 ggtttggaataattcaaggcgagagaccaaatttataactggttaaatatggcggagat 1362  
 Db 1621 GGTGACGAAAACCTACCAAGGCGGAAAAACAATTTACAATGTCACAACTATGGAGGAGAT 1680  
 Qy 1363 gacqcatattaatacagagacttatttagaattctttccaatgatgattataagagtcgg 1422  
 Db 1681 GACGCTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTT 1740  
 Qy 1423 gcaacttgggctaactcagctcgttatcctcaaacataaataaaggacgcaaaaagaatcct 1482  
 Db 1741 GCATCATGGGCAAAATGGTTCTATCGGTATCTTGAACCTTGATGATATGAGAAAAATCCA 1800  
 Qy 1483 aagtggcagagttatactcaccatttaattctgcacaaatgggtaagcgtgaattgag 1542  
 Db 1801 TCTGTAGCTGAACCTTTACACACCGCTTGTCACTAGCCACACATGTTACTGATGAAGTAGAA 1860  
 Qy 1543 cgaccaaattgtagttaaattaggttaataataattactatttccgctacccgctttaaact 1602  
 Db 1861 CGTCCAAGTGTGTTAAGATGGGTGGCAAGTATTAACCTCTTCACGTCTTACCGTATCAAC 1920  
 Qy 1603 cgagggaagtaagatgctggtggaatgtaattatgcccgttggtgataatgttgca 1662  
 Db 1921 AAATCTACGGATGCTGAAGGTACTGTAGCAGCGCGTGAAGCTGCTGCTGATGACGTTGTC 1980  
 Qy 1663 atggtcggatattgctgatagtctaaactggtattataagccattaaatgattctgga 1722  
 Db 1981 ATGCTTGGATTGTTATCAGACAGCCTCCGAGGAAAATATCGTCCACTTAATGCTCAGGT 2040  
 Qy 1723 gtagttgactgctctgttctctgcaaaactggcgacagcaacttattcatattatgct 1782  
 Db 2041 GTGGTCTTGACAGCCTCTGTACCAGCTGACTGGCTGACATCTACCTATTCTTACTATGCG 2100  
 Qy 1783 gtccccgttgcggaaaaagatgacaaagtatttagttacttcaatatagactaaatagaaat 1842  
 Db 2101 GTTCCTCTTTGAAGGCTCATCAGATACTCTTTTGGTAACTTCTTTACATGATGATACCCGTG 2160  
 Qy 1843 ggagtacggggttaaaaggaatggattcaacttggcgacagctgttcttactacaaaataac 1902  
 Db 2161 GGAATTCGCGGTGCTGAGAATAAATCAACATGGCTGCCAAGCTTCTTATCAAGATGAAT 2220  
 Qy 1903 cgggataacacaaactactgcttttagttaaataagtaatacaagggaattggatttggat 1962  
 Db 2221 GCAGATGATACACAGAAGTTCTTCTTAAGATGACCAATCAAGGTGACGTGATTTGGGAT 2280  
 Qy 1963 gattcaagcgaaaaactcttgatattgattgatttagactccgctctttaccctggcgaa 2022  
 Db 2281 AAATCTAGTGAAGCTTGGTACAGTGGCGATCAAAACAGTCTAAATTTGCCAACAGAG 2340  
 Qy 2023 cgtgataaaccttggattggggaacttaattgggttaggtataaaacccgcatgactcctgct 2082

```

Db 2341 GACITTAATGATATTATGCGGTATCGGTTATGCTTAACACACATACTATCCT 2400
Qy 2083 acacaaatgacc 2096
    ||| ||| |||
Db 2401 ACAGTTGATGCTT 2414

RESULT 7
AE007686
LOCUS
DEFINITION
    Clostridium acetobutylicum ATCC824 section 174 of 356 of the
    complete genome.
ACCESSION
    AE007686 AE001437
VERSION
    AE007686.1 GI:15024739
KEYWORDS
    SOURCE
    ORGANISM
        Clostridium acetobutylicum.
        Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
        Clostridium.
REFERENCE
    1 (bases 1 to 11318)
    Nolling,J., Breton,G., Omelchenko,M.V., Markarova,K.S., Zeng,Q.,
    Gibson,R., Lee,H.M., Dubois,J., Oiu,D., Hitti,J., Wolf,Y.I.,
    Tatusov,R.L., Sabathe,F., Doucette-Stamm,L., Soucaille,P.,
    Daly,M.J., Bennett,G.N., Koonin,E.V. and Smith,D.R.
    Genome Sequence and Comparative Analysis of the Solvent-Producing
    Bacterium Clostridium acetobutylicum
    J. Bacteriol. 183 (16), 4823-4838 (2001)
    21359325
    11466286
REFERENCE
    2 (bases 1 to 11318)
    Childress,D., Zeng,Q. and Smith,D.R.
    Direct Submission
    Submitted (24-JUL-2001) GTC Sequencing Center Production,
    Finishing, and Bioinformatics teams, Genome Therapeutics Corp., 100
    Beaver Street, Waltham, MA 02453-8443, USA
    Location/Qualifiers
    1..11318
    /organism="Clostridium acetobutylicum"
    /strain="ATCC 824"
    /db_xref="ATCC:824"
    /db_xref="taxon:1488"
    89..790
    /gene="CAC1770"
    89..790
    /gene="CAC1770"
    /codon_start=1
    /transl_table=11
    /product="RNA polymerase sigma factor, Sigk-like"
    /protein_id="AAK79735.1"
    /db_xref="GI:15024740"
    /translation="MLNRSLEDRIQAKESKEELNRLIQEYKPFASLQKTKGYL
    QYGDLSLGMATFEAVSDYRNKGKFLSFAKVIILRAIDYRNQKFNKVILO
    GNYDENNDINPMFNKAMQYKNKSGINEIRLEILEYKELREWGTEFVNLVQASP
    QELKKTYKNIANVIVHNQDILNKLIRVKRLPIKEIEKNIYINRKLGRGIYIALV
    IAEIGDYRLRIEYIE"
    801..1544
    /gene="CAC1771"
    801..1544
    /gene="CAC1771"
    /codon_start=1
    /transl_table=11
    /product="Uncharacterized protein, ykri B.subtilis
    homolog"
    /protein_id="AAK79736.1"
    /db_xref="GI:15024741"
    /translation="MKGIVIEELKYLIILTEKGFIEIDKLNTRAKIGOEINIKSKK
    ISTKALRRPALAATLILFFIINSLYGMIFTRVVDINPNKNKNVAMEIHYNF
    GNTIKQAANKGVSIVKRNRLKFKATVNNFIRTAQNDKFIDSGKQNTIITIA
    SSKSINDESIDSSLEHYIKENKINARPMIVLGNIDIDYKKSQVGIPIDKFILINKVI
    KNPSPYKFDLKKSIDELINTANQENY"
    1682..3151
    /gene="CAC1772"
    /note="sacB1"

FEATURES
    source
    gene
    CDS
    1682..3151
    /gene="CAC1772"
    /codon_start=1
    /transl_table=11
    /product="Levansucrase"
    /protein_id="AAK79737.1"
    /db_xref="GI:15024742"
    /translation="MNKLKIVKILIGSMICSGIITQOTFASTDMNMYKEYGVSHIT
    RYNMSKIPMEQNDLAFKYPQFNASTLKNIASAKGYDKNGNLIDLDVDSWPLQNGDGT
    VANYGYHIYFALAGDPKQDDTSIYMFYQKIGENSIDSWKNAGKVFDSKQVANDP
    YIKYQTOGWSGATLSDGQVRLFTDFSGVAKDGGTDSANQVITFTVNLSPDSNP
    INIDSVDHKSVDGNGETIYQVQOFIDEGKSSGDNHTLRDPHYVEDNGKYLVE
    ANTGTNDGQDGTSLNKAIFYGRSOFTEKQDILLIDTNKHDSLANGALGIIELN
    NDTLTKEMKPLIASNTVDETERANVFMNGRWLYFTDSRGSKMTINGISKDIYML
    GFSSNLGTCPYKPLANKTGLVNLNLDPTDFTFTYSHFAVPQTNGKNVVTITSYITNRGM
    YSDHHSSFAPSEFLNLNKGTKTSVINSILQOGLTIDNY"
    3249..4793
    /gene="CAC1773"
    /note="sacC"
    3249..4793
    /gene="CAC1773"
    /codon_start=1
    /transl_table=11
    /product="Levanase"
    /protein_id="AAK79738.1"
    /db_xref="GI:15024743"
    /translation="MKRKKLLTILMLFTVILVILLSLYIQNSTIKSTTTKKQNY
    REVYHFTVPNKWKNDDPQPVFDEGEHYHYLYLNKDYPNGNGTEMRQATSKDLVTWKDE
    GVCIPKTYNENGDIWSGSFVVDQAQNTAGGAGVAIVTQPSASMDQEQFLMTSTDR
    GRTEKPYSNQPIIENPCPKDPRDKIIMDFKNKMWVLAEGTKIGEYSNLKNQOH
    TGDEFTNNIGIVECPDIFPMOSNDGNTYKILGTSAANGKVSCEPTNYAYWGNVGGKF
    IADISTPKWDYGFEDWYAATFESENQSKNLKRYALAMNNNDYANNTPTIQNRFNG
    MDSVREITTSKQNDTYSLSKPIKRIENITTSIDQFKQISVYGLKHLKVGQKAYQL
    DTDINWDANKVGRILRESLIDKRRIDVGTCKEYSVNRATYTCNEDSKKYYVESRA
    PFDLNKKVHLRIEFDKVSVEVFIDDKGIYSNEVFPREDKGTILFSIKGAVFEKNI
    TIKHID"
    4860..6146
    /gene="CAC1774"
    /note="sacB2"
    4860..6146
    /gene="CAC1774"
    /codon_start=1
    /transl_table=11
    /product="Levansucrase"
    /protein_id="AAK79739.1"
    /db_xref="GI:15024744"
    /translation="MKRTKYKMLSSLMVILAILTIPFLILRHNTGYTSTWSRQAO
    FKCTENTAPNINPNFKLTAPNLWVNDTWPLVKDGLAVVNGYKILFALTARNVGM
    NKRHDVAGISYFCSTGDNVNYKGLAYNEDALGSRQWAGSAILDENGVMVOFFYATG
    RKCEAVPTFQRLVKYKFSINVDKGGVHITNCKHQVILEPDGVYIYQTMQOAGKPIY
    SFDPFTFEDQTKDYLIPEFGKGGKIEKMPENIGDKLFRKNHIAPRGVENFNGV
    GIAVANQDUTREKLLPPLLEAVGVNQQLERQIVMKKNKYILFTISHKFTYAAGLNG
    VDLYGFCGNSLRSNYKPLNGNLVITNTPTNDPYQYSWYLVSGHDVLSFINHYHFG
    QLRYGTFAPPTQISLKGYSKIIGRLGEGVWTPAH"
    6560..8305
    /gene="CAC1775"
    6560..8305
    /gene="CAC1775"
    /codon_start=1
    /transl_table=11
    /product="Predicted membrane protein"
    /protein_id="AAK79740.1"
    /db_xref="GI:15024745"
    /translation="MKIGIIAISMISILLIGSISSVVKASSNNKRLQNLTYKEFOK
    LKNSKSLVRDNDNNEKLSLDLKNKKYIPFNKTIILKNKIKLPIKSGQDSVSKY
    GNSVASAGDSNNAFPAKLDYILNPSLNGEITNTQIAIWDSDTDADGALSRA
    VGGFTSAYTLGVLDQNKGEVTOFTNAGDLDLYQVTDGSKGASNLVGSYIHPEDY
    QAQFQSTTQTDVTKYVNAVDFSKMDTAICLTGTSKDMNLKISDQSGNVQVNY
    DKKYYINKPSSNAGICNYITITPTVSSNSTYSYRLMIGNKNDKSEAMLSIGQNTI
    ILDKYTENVKQFIOEYTPKNQOEKFTADGSTDITLLNNYPELRFQVRDTEPNDL
    FTDPNNAHRSQCGNLNYKAEKARLNTAGQVYLYVTSPLIQDPTLNDIHL
    LVGDPVLLYDSTTAYADNYITGYSSTIANINVTNVPPTAAVKDIYVRSINDVAR
    VSDFNKVMAPGSFWRDGGQCCIGIDVNVYVNGTNNVQLMGTRFVFGQGVDPIS

```







```

/note="pot. rRNA binding site (sacB)"
464..550
464..1885
/codon_start=1
/transl_table=11
/product="levansucrase"
/protein_id="CAA26513.1"
/db_xref="GI:732568"
/db_xref="SWISS-PROT:P05655"
/translation="MNIRKFAQATVLTFTTALLAGGATQAFKAKETNOKPKYKYGIS
HYTHDMQIPQOKNEKYQVPEFDSSTIKNISSAKGLDWDWSPLOADGTVANVHG
YHVFALAGDPKADDTSIYMEYQKVGTSIDSWKNAGRVFKDSKPDANDSILKDOT
QWSSGATFTSDGKIRLFTYDFSGHYGKQTLTTAOVNVASDSLSLNTNGVEDYKSI
DGDGTYQNVQOFIDEGNYSSGDNHTLRDPHYVEDKGHKYLVFEANTGTDEGYGEE
LNPKAYGKSTSFQRESQKLLQSKKRTAELANGALMIENDDYTLKVKMPLIAS
NVTDEIRANVFNKRWYLFDSRSGKMTIDGTSNDIYMLGYVSNLSLTPKPLN
KTGLVLKMDLPNDVTFTYSHFAVPOAKGNVVIITSYMTNRGFYADKQSTFAPSFLN
IKGKTSVVKDSILBQGLTVNK"
551..1882
/mat_peptide
/product="mature Levansucrase (aa 1-444)"
1892..1904
/repeat_unit
/note="inverted repeat B"
1892..1928
/misc_feature
/note="region of dyad symmetry"
1916..1928
/repeat_unit
/note="inverted repeat B"
BASE COUNT 720 a 422 c 370 g 495 t
ORIGIN

```

Query Match 3.4%; Score 80; DB 1; Length 2007;  
Best Local Similarity 47.7%; Pred. No. 3.3e-05;  
Matches 464; Conservative 0; Mismatches 460; Indels 48; Gaps 6;

```

Qy 983 cacaagaatggtcagagatcagctgtttgaacagtgataactctatccaattatttata 1042
Db 942 CACAAGATGGTCAGGTCAGCCACATTACATCTTGAGCGAAAAATCCGTTTATTCTACA 1001
Qy 1043 caagggtgacagctctgtatacaataccaatacatcaaaaaattgctagcactcttt 1102
Db 1002 C-----TGATTTCCCGTAAACATTACGGCAACAACACTGCAACTGCACAAAGTTA 1055
Qy 1103 atttaactgataataatggaatgtatcactcgtcaggtacgaatgactattgtat 1162
Db 1056 AGTATCAGCATCAGACAGCTCTTTGAAACATCAACGGGTAGAGGATTATAAATCAATCT 1115
Qy 1163 ttgaaggtgagctattactaccaaaacttatgatcaatgga-----aagctactaacaag 1219
Db 1116 TTGACGGTGACGGAAAAACGTATCAAAATGTACAGCATCTCATCGATGAAGCAACTACA 1175
Qy 1220 gtcgcgataatattgcaatgctgctcatgtaattgaagatggtaatgggtgatcggt 1279
Db 1176 GCTCAGGCGACAACCATACGCTGAGAGATCTCTACTAGTAGAGATTAAGGCCACAAT 1235
Qy 1280 acctgttttgaagcaagtactgtttgaaaaattatcaaggcaggaacaaatttata 1339
Db 1236 ACTTAGTATTTGAAGCAAAACCTGGAACCTGAAGTGGCTACCAAGGCGAAGATCTTTAT 1295
Qy 1340 actggttaaatattgcgagatgacgcatttaataatacaagcttatttagaattttt 1399
Db 1296 TTAACAAGCATATATGGCAAAAGCACATCTTCTCCGTCAAGAAGTCAAAAACCTC 1355
Qy 1400 ccaatgatgatatcaagagtcggaactggcgctaatgacgctatcgttatcctcaaac 1459
Db 1356 TGCAAAAGCATATAAAACGACGCGCTGAGTTAGCAACGCGCTCTCGGTATGATTCAGC 1415
Qy 1460 taaataggacgaagaagaatcctaagtggcagaggttatactcaccaattatttctgcac 1519
Db 1416 TAAACGATGATTACACACTGAAAANAAGTATGAA-----ACCCTGATTGCATCTA 1466
Qy 1520 caatggtaagcgtatgaattgagcgacccaatgtagtttaattagggttaataataattact 1579
Db 1467 ACACAGTAACAGATGAATTTGAACGCGCAACGCTCTTTAAATGACGCGCAATGGTACC 1526

```

```

Qy 1580 tattgcgcgtaccggttttaaatcagaggagtaagttaagtatgctgttggaatgctaat 1639
Db 1527 TGTTCACGTACCCCGCGGATCAAAAATGACGATTGACGCGCATTTAGCTTAACGATATTT 1586
Qy 1640 atgcggttggtgataaatgtgcaatggtcggtatgtgctgctgctgctgctgctgct 1699
Db 1587 A-----CATGCTTGGTTATGTTCTTAATTTCTTAACTGCCCAT 1625
Qy 1700 ataaagccattaaatgattcgtgagtagtgctgctgctgctgctgctgctgctgctgct 1759
Db 1626 ACAAGCCGCTGAACAAAACGTGCCTTGTTGTTAAAAATGGATCTTGATCTCAACGATGAA 1685
Qy 1760 cagcaactattcatatcatctgctcccggttgcgcggaagaaagatgacccaagtattagta 1819
Db 1686 CTTTACTTACTCACATCTCGCTGTACC---TCAAGCGAAGGAAACAATGTCGTGATTA 1742
Qy 1820 ctctatatgactaatagaaatggagtagcggtgtaaaggaaatgattcaactgggcac 1879
Db 1743 CAAGCTATATGACAAACAGAGGATTCTACGCAGACAAA-----CAATCAAGTTTGGC 1796
Qy 1880 cgagtttcttactcaaaaataacccgggatacaacaaactactgttttagctaaaaagta 1939
Db 1797 CAAGCTTCTGCTGAACATCAAAAGGCAAAACATCTGTTGTCAAAGACAGCATCCTTG 1856
Qy 1940 atcaagggtgatt 1951
Db 1857 AACAGGACAAAT 1868

```

```

RESULT 10
SYNPRC/c 5941 bp DNA circular SYN 27-APR-1993
LOCUS
DEFINITION Cloning vector (pRL278) for sacB-mediated positive selection for
double recombinants in gram-negative bacteria.
ACCESSION L05083
VERSION L05083.1 GI:209132
KEYWORDS Cloning vector.
SOURCE Cloning vector DNA.
ORGANISM unidentified cloning vector
artificial sequence; vectors.
REFERENCE 1 (bases 1 to 5941)
AUTHORS Cai, Y.
TITLE Molecular genetic approaches towards the understanding of
heterocyst differentiation and pattern formation in the
cynobacterium Anabaena sp
JOURNAL Unpublished (1992)
REFERENCE 2 (sites)
AUTHORS Black, T.A., Cai, Y. and Wolk, C.P.
TITLE Spatial expression and autoregulation of hetR, a gene involved in
the control of heterocyst development in Anabaena
JOURNAL Unpublished (1992)
FEATURES
source 1. .5941
/organism="unidentified cloning vector"
/db_xref="taxon:45196"
complement(473..1894)
/note="sacB from B.subtilis; ORF"
/codon_start=1
/transl_table=11
/protein_id="AA72307.1"
/db_xref="GI:209133"
/translation="MNIRKFAQATVLTFTTALLAGGATQAFKAKETNOKPKYKYGIS
HYTHDMQIPQOKNEKYQVPEFDSSTIKNISSAKGLDWDWSPLOADGTVANVHG
YHVFALAGDPKADDTSIYMEYQKVGTSIDSWKNAGRVFKDSKPDANDSILKDOT
QWSSGATFTSDGKIRLFTYDFSGHYGKQTLTTAOVNVASDSLSLNTNGVEDYKSI
DGDGTYQNVQOFIDEGNYSSGDNHTLRDPHYVEDKGHKYLVFEANTGTDEGYGEE
LNPKAYGKSTSFQRESQKLLQSKKRTAELANGALMIENDDYTLKVKMPLIAS
NVTDEIRANVFNKRWYLFDSRSGKMTIDGTSNDIYMLGYVSNLSLTPKPLN
KTGLVLKMDLPNDVTFTYSHFAVPOAKGNVVIITSYMTNRGFYADKQSTFAPSFLN
IKGKTSVVKDSILBQGLTVNK"
2749..3543
/note="kanamycin and neomycin resistance determinant"
CDS

```



```

(npt11)"
/codon_start=1
/transl_table=11
/protein_id="AAA72308.1"
/db_xref="GI:209134"
/translation="MTEODLHAGSPAAWVERLFGYDWAQOTIGCSDAAVFRLSAGR
PVLVKTDLSSGALNELQDEARLSWLATGVCACAVLDVYVTEAGRWMLLGEVPGDL
LSHLAPAEKVSIMADMRLLHTLDPATCFDHOAKHRIERAKTRMEAGLVQDDILDE
EHQGLAPAEFLARKAMPEDGLVTHGDACLPLNTWVNGRFSFIDGRLGVADRY
QDIALATROIAGELGEWADRLVLYGIAAPDSQRIATFYRLLEDF"
misc_feature
3558..4815
/notes="multiple cloning sites"
misc_feature
4816..5941
/notes="oriV and oriT genes from Plasmid pMB1"
BASE COUNT      1419 a   1329 c   1471 g   1722 t
ORIGIN

Query Match      3.4%; Score 80; DB 12; Length 5941;
Best Local Similarity 47.7%; Pred. No. 2.9e-05;
Matches 464; Conservative 0; Mismatches 460; Indels 48; Gaps 6;

QY  983 cacaagaatggtcagatcagctgttttgaaacagtgataactctatccaatattttata 1042
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  1416 CACAAGAAATGGTCAGGTCAGCCACATTTACATCTGACGGAATAAAATCCGTTTATTCTACA 1357
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  1043 caagggtagacagctctgtgatacaatcaccatcacaataatgataactgataactattgtat 1162
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  1356 C-----TGATTTCCTCGGTAAACATTAGCGCAACAAACACATGACAACTGCACAAAGTTA 1303
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  1103 atttaactgataaataatgaaatgtatcactcgtcagtgatcagaataactattgtat 1162
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  1302 AGTATCAGCATCAGACAGCTCTTTGAACATCAACGGTGTAGAGGATTTATAATCAATCT 1243
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  1163 ttgaagggtatggtctattactaccacaaacttatgataatgga---aagctactaacaag 1219
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  1242 TTGACGGTGCAGGAAACGATCAATATGACAGAGTTTCATCGATGAAGCAACTACA 1183
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  1220 gtgcgataaatttgcattgctgatgctcatgtaattgaagatggttaattggtgacggt 1279
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  1182 GTCAGGCGACACCATACCGTGTAGAGATCTCTACATAGTACAGATTAAGAGCCCAAT 1123
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  1280 acctgtttttgaagcaagtagctgttttgaaaaattatcaaggcgaggacaaatttata 1339
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  1122 ACTTAGTATTGAGCAACACACTGGAACAGATGAGATGGCTACCAAGCGGAAGAACTTTAT 1063
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  1340 actggttaaatatggtgcgagatgcgcgtatcaatatacaagagctatttagaattcttt 1399
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  1062 TTAACAAAGCATACTATATGGCAAAAGCACATCATCTTCCGTCAAGAAAGTCAAAACTTC 1003
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  1400 ccaatgatgatattgaaggtcggcaacttgggtaattgaagctatcggtatcctcctcaac 1459
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  1002 TGCAAGCGATTAACAAAGCGACGCTGAGTTAGCAAGCGGCGCTCGGTATGATTGAGC 943
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  1460 taataaggaagcaaaaataactaaggtggcagagttatactaccatttaatttctcac 1519
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  942 TAAACGATGATTACACACTGAAAAAAGTGATGAA-----ACCGCTGATTGCATCTA 892
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  1520 caatgttaagcgatgaattgagcgacaaatgtagttaattggttaataataattact 1579
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  891 ACACAGTAACAGATGAATTTAGACGCGCAACGCTTTTAAATGAACGGCAATGGTACC 832
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  1580 tattgcgcgtaccctgttataatcaggaagatgatgatgctgtgataatgctaat 1639
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  831 TGTTCACGTACTCCCGGGATCAAAATGACGATTGACGGCATTACGCTCAACGATATTT 772
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  1640 atgcccgttgataatgttgcgaatggcggatattgttgcgtgatagctgaactggactt 1699
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  771 A-----CATGCTGCTGTTATGTTTCTTAATCTTTAACTGGCCCAT 733
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  1700 ataagccattaaatgattcgtgagtagctctgactgcttcttctcctgcaaacggcga 1759
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  732 ACRAAGCGCTGACAAACACTGGCCTTGTGTAAAAATGGATCTTGATCCTTAACGATGTA 673
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

QY  1760 cagcaacttattcatattatgtgtcccttgcggaaagatgacaaagtattagta 1819
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  672 CTTTACTTACTCACACTTCCTGTACC---TCAAGCGAAAGGAAACAATGTGCTGATTA 616
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  1820 ctctatatgactaatagaaatggagtagcggtgtaaaaggaatggatcacttggcgcac 1879
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  615 CAAGCTATATGACAAACAGAGGATTTACGCGACACAAA-----CAATCAACGCTTTGCCG 562
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  1880 cgggtttcttactacaaataaccccgataacacactactgttttagctcaaatgacta 1939
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  561 CAAGCTTCTCTGCTGAACATCAAAAGCGCAAGAAACATCTGTTGCAAGACAGCATCCTTG 502
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  1940 atcaaggggatt 1951
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  501 ACAAGGACAAAT 490
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 11
SYNPLA/c
LOCUS
DEFINITION
SYNPLA 6347 bp DNA circular SYN 27-APR-1993
Cloning vector (pRL271) for sacB-mediated positive selection for
double recombinants in gram-negative bacteria.
ACCESSION
L05081
VERSION
L05081.1 GI:209125
KEYWORDS
cloning vector DNA.
SOURCE
unidentified cloning vector
artificial sequence; vectors.
ORGANISM
1 (bases 1 to 6347)
REFERENCE
1. Cai, Y.
AUTHORS
Cai, Y.
TITLE
Molecular genetic approaches towards the understanding of
heterocyst differentiation and pattern formation in the
cynobacterium Anabaena sp
JOURNAL
Unpublished (1992)
AUTHORS
2 (sites)
Black, T.A., Cai, Y. and Wolk, C.P.
TITLE
Spatial expression and autoregulation of hetR, a gene involved in
the control of heterocyst development in Anabaena
JOURNAL
Unpublished (1992)
FEATURES
Location/Qualifiers
source
1..6347
/organism="unidentified cloning vector"
/db_xref="taxon:45196"
complement(473..1894)
/notes="sacB gene from B.subtilis; ORF"
/codon_start=1
/transl_table=11
/protein_id="AAA72302.1"
/db_xref="GI:209126"
/translation="MNIKFAKATVLTFTTLAGGATQAFAKETNOKPYKETYGIS
HITRDMQLIPEQOKNEKYOVPEFDSSTIKNISAKGLEWDSWPLQNDGTVANYHG
YHIVFALQDPKNADDTIYFVKVGETSIDSKNAGRVFKDSDFDANDSTILKDOT
QEWGSGATFTSDGKIRLFTYDFSGKHVGTQLTAAOVNYSASDSSININGVEDYKSF
DGDGKTYQNYQQFIDEGNTSSGDNHTLRDPHYVEDKHLYLVEANTGTEDGTQGEES
LNFKAYGKSTSPFROESOKLLOSDKRTAELANGALGMIELNDDITLKKVMKPLIAS
NTVTDIERANVFKNMKWYLFPTDSRGSKMTIDGITSNDIYMLGYVNSLTGPKPLN
KTCGLVKMLDLPNDVFTTSHFAPQAGKNNVVITSYMTNIRGFADKQSTFAPSFLN
IKGKTSVVRKDSILEOGLTVNK"
complement(2393..3127)
/notes="erythromycin resistance determinant"
/codon_start=1
/transl_table=11
/protein_id="AAA72303.1"
/db_xref="GI:209127"
/translation="MNEKNKHSQNPITSKHNIDKIMTIRLNEHNDIFEIGSGKGHF
TLELVKRCNEVTAIEIDHKLTENKLVHDHNFVNLKDILOFEPKQNSKYKIGNI
PYNTSDIIRKIVFDSIANEYILVEYGFARLLATKRSALLLAEVDISLISVPR
EYFHPKPVNSSLIRLSRKKSIRSHDKQKYNFVWKWVKNYKFKIKFTKNFNSLKH
AGIDDLNNTSFEQFLSLFNSYKLPNK"
3503..4162
/notes="chloramphenicol resistance determinant"
/codon_start=1

```

/transl\_table=11  
/protein\_id="AA72304.1"  
/db\_xref="GI:209128"  
/translation="MEKINGTVDISOWHKEFEAFOSVAGCTYNQTVOLDITAF  
LKVKNKHKFPYAFIHLARLNAPHEFMKMDGELVINDSVHPCYTVFHEQETEF  
SSUWSEYHDFRGLHIYSQDVACYGLENLAIFPKGETENNFPVSAHPWSVTFSDLN  
ANMNFDFAPVFTMGKITYTQDKVLMPLAIOVHVHVCDFHVGRLNELAQYCDWOGG  
A"

misc\_feature 4252..5221  
/note="multiple cloning sites"  
misc\_feature 5222..6347  
/note="oriV and oriT genes from Plasmid pMB1"

BASE COUNT 1649 a 1305 c 1404 g 1989 t  
ORIGIN

Query Match 3.4%; Score 80; DB 12; Length 6347;  
Best Local Similarity 47.7%; Pred. No. 2.8e-05;  
Matches 464; Conservative 0; Mismatches 460; Indels 48; Gaps 6;

Qy 983 cacaagaatggtcaggtcagctgttttgaaacagtgataaactctctatccaattattttata 1042  
|||||  
Db 1416 CACAAGAATGCTCAGGTTTCAGGCACATTACATCTGACGGAAAAATCCGTTTATTCTTACA 1357  
|||||  
Qy 1043 caagggtagacagctgataacaataccaatcatcaaaaaatgtagcgctactcttt 1102  
|||||  
Db 1356 C-----TGATTCTCCGGTAAACATTACCGGCAAAACAACTGACAACTGCACAAAGTTA 1303  
|||||  
Qy 1103 atttaactgataaatggaatgtatcactcgctcaggtcacgaatgactatattgtat 1162  
|||||  
Db 1302 ACGTATCAGCATCAGACAGCTCTTTGAACATCAACGGTGTAGAGGATTATAATCAATCT 1243  
|||||  
Qy 1163 ttgaagtgatgcttattactacaaacttatgatcaatgga---aagctactaaacaag 1219  
|||||  
Db 1242 TTGACGGTGCAGGAAAAACGATATCAAAATGTACAGCAGTTTCATCGATGAAGGCAACTACA 1183  
|||||  
Qy 1220 gtgcgataataattgcaatgctgagctcatgtaattgaagatgataatgtagtcggt 1279  
|||||  
Db 1182 GCTCAGCGCACACCATACGCTGAGAGATCCTCAGTACGTAGAGATGAAGGCGCAAAAT 1123  
|||||  
Qy 1280 acctgttttgaagcaagtactggtttggaataattcaagcgagggacaaattata 1339  
|||||  
Db 1122 ACTTAGTATTGAAGCAAAACACTGGAAGTCAAGATGCTACCAAGCGCAAGATCTTTAT 1063  
|||||  
Qy 1340 actggttaattatggtcagagatgcacatttaataatcaagagcttatttagaattcttt 1399  
|||||  
Db 1062 TTAACAAGCATACTATGCGAAGAACACATCATCTTCCTCGTCAAGAAAGTCAAAAACCTTC 1003  
|||||  
Qy 1400 ccaatgatgatattaagagtcgggcaacttgggctaatgcagctatcggtatcctctcaaa 1459  
|||||  
Db 1002 TGCMAAGCGATAAAAAACGACGGCTGAGTTAGCAACGCGCTCTCGGTATGATTGAGC 943  
|||||  
Qy 1460 taataaggaacgaagaatcctaaggtggcagagttatactaccatattattctgcac 1519  
|||||  
Db 942 TAAACGATGATTACACACTGAAAAAAGATGATGAA-----ACGCGTATTGTCATCTA 892  
|||||  
Qy 1520 caatggtgaagcgtgaaattgagcgacaaatgtagttaattaggtaataataattact 1579  
|||||  
Db 891 ACACAGTAACAGATGAATTAAGCGCGCAAGCTCTTTAAATGAACGCGCAATGGTACC 832  
|||||  
Qy 1580 tatttgcgctaccctgttaaatcaggaagtaatgatgcttggatgataatgcttaatt 1639  
|||||  
Db 831 TGTTCACTGACTCCCGCGGATCAAAAATGACGATTGACGGCATTAAGTCAACGATATTT 772  
|||||  
Qy 1640 atgcggttggataatgttgcattggtcgatattgctgatatgctataactgatacttt 1699  
|||||  
Db 771 A-----CATGCTGGTATGTTTCTTAATCTTTAACTGGCCCAT 733  
|||||  
Qy 1700 ataagccataaatgattctggagtagtcttgcagcttctgttctgcgaactggcgga 1759  
|||||  
Db 732 ACAAGCGCTGAACAAAACACTGGCCTTGTTTAAAAATGGATCTTGATCTTAACGATGTA 673  
|||||  
Qy 1760 cagcaacttaticatattatgctgctcccctgttgcgcggaagaagatgacaaagtattagta 1819  
|||||

Db 672 CCCTTACTTACTCAGCTTCCTGTACC---TCAAGGAAAGGAAACAATGTGCTGATTA 616  
|||||  
Qy 1820 cttcatatgactaatagaaatggagtagcggtggaagaaatggattcaacttggcgcac 1879  
|||||  
Db 615 CAAGCTATATGACAAACAGAGGATTCTACGACACAAA-----CAATCAACGTTTGGCG 562  
|||||  
Qy 1880 cgaattcttactacaaatgaaccgggataaacacactactgttttagctaaatgacta 1939  
|||||  
Db 561 CAAGCTTCTGCTGAACATCAAGGCAAGAAACATCTGTTGTCAAAGACAGATCCTTG 502  
|||||  
Qy 1940 atcaaggggatt 1951  
|||||  
Db 501 AACAAAGGACAAAT 490  
|||||  
RESULT 12  
SYNPRLB/c 6798 bp DNA circular SYN 27-APR-1993  
LOCUS  
DEFINITION SYNPRLB Cloning vector (pRL277) for sacB-mediated positive selection for double recombinants in gram-negative bacteria.  
ACCESSION L05082  
VERSION L05082.1 GI:209129  
KEYWORDS cloning vector.  
SOURCE Cloning vector DNA.  
ORGANISM unidentified cloning vector  
REFERENCE 1 (bases 1 to 6798)  
AUTHORS Cai,Y.  
TITLE Molecular genetic approaches towards the understanding of heterocyst differentiation and pattern formation in the cyanobacterium Anabaena sp  
JOURNAL Unpublished (1992)  
AUTHORS Black,T.A., Cai,Y. and Wolk,C.P.  
TITLE Spatial expression and autoregulation of hetR, a gene involved in the control of heterocyst development in Anabaena  
JOURNAL Unpublished (1992)  
FEATURES  
source 1. 6798  
/organism="unidentified cloning vector"  
/db\_xref="taxon:45196"  
complement(473..1894)  
/note="sacB gene from B.subtilis; ORF"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AA72305.1"  
/db\_xref="GI:209130"  
/translation="MNKIFAKQATVLTFTTALLAGGATQAFKTNOKPYKETYGIS  
HITRDMIQIPEQOKNEKYQPEFDSSTIKNISAKGLEWDSWPLQNDGTVANYHG  
YHIFALAGDFKNADTSIYMFYQKVGSTSDSKNAGRVKDSDFDANDSILKDOT  
OENSGSATFTSDGKIRLFTYDFSGKHGKQKLTAAOVNVSASDSSLNNGVEDYKSF  
DGDGKTYNQVOFTDEGNYSSGDNDHTRDPHYVEDKHGKHLVFEANTGCTEDGYGEE  
LFNKAYYKSTSFPROBOKLLOSDKRTAELANGALHIELNDYTLKVKVKKPLIN  
NTVDELERANVFNMKGWKLFTDSRGSMTDITGTSNDIYMLGTVSNLSLTPKPLN  
KTGLVKMDLPNDVTYSHFAVPOAKGNHVITSYMTNRGFAKQSTPAPSPFLN  
IKGKTSVVKDSILEQGLTVNK"  
complement(2980..3771)  
/note="streptomycin and spectinomycin resistance  
determinant (aadA)"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AA72306.1"  
/db\_xref="GI:209131"  
/translation="MREAVIAEVSQTSLEVVGVIERHLEPTLLAVHLYGSAGVGLKLP  
HSDLDLVTVTRDLDETTRALINDLETSPGSEILRAVEVTIVHDDIIPWYVP  
AKRELOFGWQRNDILAGIFEPATIDIDLALITKAREHSVALGPAARELFDVBPQ  
DLFALNETLANSPPDWDAGDERNVVLTLSRIWYSATGKIAPKQVADWAMERLPA  
QYQPVILEARQATLGQEDRLASRADQLEEFVHYVKGTEITVKGK"  
4241..5504  
/note="multiple cloning sites"  
5667..6798  
/note="oriV and oriT genes from Plasmid pMB1"  
misc\_feature  
misc\_feature

```

BASE COUNT      1653 a      1540 c      1637 g      1968 t
ORIGIN

Query Match      3.4%; Score 80; DB 12; Length 6798;
Best Local Similarity 47.7%; Pred. No. 2.8e-05;
Matches 464; Conservative 0; Mismatches 460; Indels 48; Gaps 6;

QY 983 cacaagaatggtcagatcagctgtttgaacagtgataactctatcccaattatttata 1042
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1416 CACAAGAATGTCAGGTCACGCCACATTTACATCTGACGGAAAAATCCGTTTATTCTACA 1357
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1043 caagggtagacagctgatacaataaccaatcatcaaaaaattgctagcgtactcttt 1102
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1356 C-----TGATTTCCTCGGTAAACATTTACGGCAAAACAACTGACAACTGCACAAGTTA 1303
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1103 atttaactgataataatgaaatgtatcactgctcaggtacgaaatgactatattgtat 1162
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1302 ACGTATCAGCATCAGACAGCTCTTTCAACATCAACGGTGTAGAGGATTTAAATCAATCT 1243
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1163 ttgaaggtgactgctattactacaaactatgatacaatgga---aagctactacaaga 1219
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1242 TTGACGGTACGGAAACAGTATCAAAATGTACAGCATCTCATCGATGAAGCAACTACA 1183
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1220 gtgcccgaataattgcaatgcgtgatgctcatgttaattgaagatggttaatgctgctg 1279
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1182 GCTCAGCGGACAAACCATACGCTGAGAGATCCTCACTACGTAGAAGATAAGGCCACAAT 1123
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1280 acctgtttttgaagaagtaactggtttggaaaaatatcaaggcagagaccataattata 1339
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1122 ACTTAGTATTTGAAGCAAAACACTGGAACCTGAAGATGGCTTACCAAGCGCAAGAACTTTAT 1063
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1340 actggttaattatgcccagagatgcgcatttaataatcaagagcttattagaattcttt 1399
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1062 TTAACAAAGCAATACATATGCGRAAAGCAATCATCTTTCGTCAGAAAGTCAAAAACCTTC 1003
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1400 ccaatgatgataatgaagatgggcaacttgggctaaatgcagctatcggtatccctcaaac 1459
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1002 TCGAAAGCGGATAAAAACGCGAGCTGTAGTATGCAAAAGCGGCTCTCGGTATGATTGAGC 943
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1460 taaataagcagaaaaagaatcctaaggtggcagagttatactcaccataattctgcac 1519
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 942 TAAACGATGATTACACATCTGAATAAAGTATGAA-----ACCGTGATTGTCATCTA 892
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1520 caatggtgaagcagataatgagcagcaaatgttagttaaataggtaataataattact 1579
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 891 ACACAGTAAACAGATGAATTTGAACGCGGAACGCTCTTTAAATGAACGGCAATGGGTACC 832
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1580 tatttgcgcctaccgctttaaactcaggaagtaatgatgatgcttggatgaatgctaatt 1639
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 831 TGTTCACTGACTCCGCGGATCAAAAATGACGATTGACGGCATATTACGCTTAACGATATTT 772
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1640 atgcccgttgggtgataatgttgcgaatggttgcgtgatgttgcgtactgactgactct 1699
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 771 A-----CATGCTGTTGTTATGTTCTTAATCTTAACTGGCCCAT 733
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1700 ataagcattaaatgattctggagtagtcttggactgcttcttgcgaactggcgga 1759
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 732 ACAAGCCGCTGAACAAACATGGCCCTTGTGTAAATAATGGATCTTGATCCTTAACGATGTA 673
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1760 cagcaactattcatattatgtcttcccgttgcggaaagaagatgaccaagtattagta 1819
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 672 CTTTACTTACTCACATCTCGCTGTACC---TCNAGCGAAAGGAANAATGTCGTGATTA 616
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1820 ctccatataactagataaaggagtagcgggtaaggaaatggattccaacttgggcac 1879
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 615 CAAGCTATATGACAAACAGAGGATTTTACGCAGACAAA-----CAATCAACGTTTGGCG 562
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1880 cgaatttctactacaataatcccggaataacacaactactgttttagctaaaatgacta 1939
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 561 CAAGCTTCTGCTGAACATCAAGGCAAGAAACATCTGTTGTCAAAGACAGCATCCTTG 502
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1940 atcaaggggatt 1951
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```





QY	1880	cgagtttctactacaaataaccgcgataacacacactactgttttagctaaatgacta	1939
Db	44816	CAAGCTTCCTGCTGAACATCAAGGCAAGAAACATCTGTTGTCAAGACAGCATCTTG	44757
QY	1940	atcaagggatt	1951
Db	44756	ACAAGGACAT	44745
RESULT	15		
BSUB0018			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
FEATURES			
source			
gene			
terminator			

```

/codon_start=1
/transl_table=11
/protein_id="CAB15324.1"
/db_xref="GI:2635832"
/db_xref="SPTREMBL:O34782"
/transl_table="MDOLKEKLVLTGSTSGIGKAAKSFLOEAGRAVYNGRKOETV
DRTIEELSGVGVHGAADLKSDTEAAAFIEKNEIGDIDILYNNGRKPOFADVT
DEMNQYFEVNVMSAVRTPHPLFKMLAKNSGRILNIAEAGVKPLPTMIPKFSMTA
LISLRGMAEMTKGNTVNSVLPGETWTEGVASYMEGAQAAGQDTDFIKDYFRVN
EPTSLQRTATAEAVNTIVFLASDAASAINGTQAVRVEGGIIRSL"
5145..5165
/terminator
/feature="yvrD"
complement(5172..6050)
/feature="yvrE"
complement(5172..6050)
/feature="yvrE"
/function="unknown"
/feature="similar to RNA polymerase"
/codon_start=1
/transl_table=11
/protein_id="CAB15325.1"
/db_xref="GI:2635833"
/db_xref="SWISS-PROT:O34940"
/transl_table="MDAVLEADTRAVIGEGPLWDEENGRLYWVDILGSELHIFDPDEK
INRSIKFSEVFTALAKYSKDELIMTKDGFYHLRDDSLKTKPKDMHESLRPNDA
KCDPYRLWAGTDSMEGEQOASLYRLNLDGSLVTKDOVSTSNGLDWRERLNYII
DPTORIVRYSDPOGSDVSNPEPVYRPQSDGLPGMTIDQNGMLWALFPGSGRYYH
IDPQKKEINSLVPAKYVTCACAFGRDLTKLITTTATQMTKEKERYEPHAGGLFSA
QLETGGYQVPVPGDV"
complement(6122..7843)
/feature="yvrG"
complement(6122..7843)
/feature="yvrG"
/function="unknown"
/feature="similar to two-component sensor histidine kinase
[yvrH]"
/codon_start=1
/transl_table=11
/protein_id="CAB15326.1"
/db_xref="GI:2635834"
/db_xref="SPTREMBL:O34989"
/transl_table="NPFHFGQMLIVLLTVMLVAFVFLDARFSDAESNGLTKATT
DTLEYLDVNDGTEVEFNFLKSVKQHWMOIIDSEGNIDYSGVPRKDPVPTTK
ELLSYTKKHLNYKLYNWAINEKSYLLSGWSKSEQLLSVEKRPQKIDSIAHY
KSSTIDYTKRKGATYLLDSNGKILDSINSTKSRKTMQLELLKYSKPNWKRIS
VKLNKDRMVAIVPNVYVTDQEFNKSLFVVLKAMFLVMAVLEMYLIIMTVWYMR
FGLPIFHTIRLVNLNLEKLEPRNREGRPVSKKKGKIKOPYRFEGLFESMDQLE
TLRRDRNREKIQATREMIAGLSHDLKPLSLIYGYSMLESKQYDWSPEEGEMGO
VYREKSEVMSKLIEDNLTYRLKNDALFIERLTSLIFFKNVIEDFKNPFSEGYDI
SFVSKEEHIEFALDARFRILNENLGNVXHNKGTEIQVILEQTKNHSILKVDNG
KQMDDEETIHLFNRYRTNTKDSGTAGTGLGLAIKALVHLNHTIHVNSRTNIGTVI
TILFKKQ"
complement(7861..8970)
/feature="yvrH"
complement(7861..8970)
/feature="yvrH"
/function="unknown"
/feature="similar to two-component response regulator [yvrG]"
/codon_start=1
/transl_table=11
/protein_id="CAB15327.1"
/db_xref="GI:2635835"
/db_xref="SPTREMBL:P94504"
/transl_table="MDGQFEQKKQKQKDEYDIEHLIACFSPIRKKLSNTSYQEREDL
EQELKIRKFEKADMLLCQDVPFGEFELYMVDENSIIISHMKKHQKQKALVCSFPAVL

```

Query Match 3.4%; Score 80; DB 1; Length 209510;  
 Best Local Similarity 47.7%; Pred No. 1.9e-05;  
 Matches 464; Conservative 0; Mismatches 460; Indels 48; Gaps 6;

QY 983 ccaagaatggtcaggatcagctgtttgaacagcagtgataactctatccaatattttata 1042  
 |||||  
 DB 136000 CACAAGAATGGTCAGCCACATTACATCTGACGCGAAATCCGTTTATTCTACA 136059

Search completed: September 22, 2002, 05:23:09  
 Job time: 20769 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2002, 01:51:26 ; Search time 7643.51 Seconds  
(without alignments)  
8182.745 Million cell updates/sec

Title: US-09-995-587A-10  
Perfect score: 4634  
Sequence: 1 gttacaaagacaaaatttt.....gttactgcttgcacaaagctt 4634

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estlin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_hic.\*  
9: gb\_esti.\*  
10: gb\_est2.\*  
11: gb\_hic.\*  
12: gb\_gss.\*  
13: em\_gss\_hum.\*  
14: em\_gss\_inv.\*  
15: em\_gss\_pln.\*  
16: em\_gss\_vit.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	106	2.3	641	12	AQ946120 Sheared D
C 2	97	2.1	700	12	AQ940248 Sheared D
C 3	92.8	2.0	687	12	BH401326 AG-ND-155
C 4	91.4	2.0	806	12	AL281759 Tetraodon
C 5	89.8	1.9	832	12	BH391984 AG-ND-138
C 6	89.4	1.9	729	12	AQ945618 Sheared D
C 7	88.6	1.8	469	9	AU076358 AG-ND-138
C 8	84.6	1.8	401	12	AQ945618 Sheared D
C 9	81.2	1.8	621	12	AQ941683 Sheared D
C 10	81	1.7	500	9	AU086536 Sheared D
C 11	81	1.7	773	12	BH391959 AG-ND-180
C 12	80.2	1.7	520	12	AQ956286 Sheared D
C 13	79.4	1.7	541	12	AL295309 Tetraodon
C 14	79.4	1.7	639	12	AQ956218 Sheared D
C 15	79.2	1.7	560	12	AQ958275 Sheared D
C 16	78.2	1.7	713	12	AZ217550 Sheared D
C 17	78	1.7	879	12	AL147405 Anopheles

18	77.4	1.7	252	9	AU052594
19	77.4	1.7	540	9	AU037631
20	76.4	1.6	584	10	BM163760
21	76.4	1.6	1007	12	BM163760 EST566283
C 22	76.2	1.6	660	12	AZ212542 T3 end of
C 23	76	1.6	498	10	AZ212542 Sheared D
C 24	75.4	1.6	519	10	BM276013 PEST0aa6
C 25	75.4	1.6	579	9	BI814949 PEST0aa0
C 26	75.4	1.6	781	12	AQ952743 Sheared D
C 27	75.2	1.6	1025	12	AL104216 Drosophila
C 28	75	1.6	653	12	AQ311782 Drosophila
C 29	74.4	1.6	522	10	BI815912 PEST0aa3
C 30	74.4	1.6	1101	12	BI815912 PEST0aa3
C 31	74.2	1.6	305	9	AU038805
C 32	73.8	1.6	471	12	BH391050
C 33	73.8	1.6	687	12	AG145469 Pan trogl
C 34	73.6	1.6	660	9	AU038978
C 35	73.2	1.6	579	9	AU038185
C 36	73.2	1.6	743	12	AZ196612 SP_1032_A
C 37	73	1.6	553	12	BH368264 AG-ND-157
C 38	72.4	1.6	522	12	BH369425 AG-ND-125
C 39	72.4	1.6	581	12	AQ658242 Sheared D
C 40	72.4	1.6	950	9	AU175513
C 41	72.2	1.6	410	9	AU037570
C 42	72.2	1.6	489	12	BH444866 BOGHF09TR
C 43	71.6	1.5	907	12	CNS021J4
C 44	71.2	1.5	500	9	AU087262 Tetraodon
C 45	71.2	1.5	571	10	BM274158 PEST0aa6

## ALIGNMENTS

RESULT 1  
LOCUS AQ946120/c  
DEFINITION Sheared DNA-46J23.TR Sheared DNA Trypanosoma brucei genomic clone  
ACCESSION AQ946120  
VERSION AQ946120.1 GI:6769385  
KEYWORDS GSS.  
SOURCE Trypanosoma brucei.  
ORGANISM Trypanosoma brucei  
REFERENCE 1 (bases 1 to 641)  
AUTHORS El-Sayed, N., Zhao, S., Gill, S., Suh, E., Malek, J., Fujii, C., Gerrard, C., Leech, V., de Jong, P., Ull, E., Melville, S., Donelson, J., Fraser, C. and Adams, M.  
TITLE Determination of clone end sequences from Trypanosoma brucei GUTat 10.1 sheared DNA library  
JOURNAL Unpublished (1999)  
COMMENT Other-GSSs: Sheared DNA-46J23.TF  
Contact: Najib M. El-Sayed  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: nelsayed@tigr.org  
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through ATCC. Sheared DNA end sequences search page: <http://www.tigr.org/tldb/mdb/tbdb/>.  
Seq primer: M13-Reverse  
Class: shotgun.

## FEATURES

Location/Qualifiers  
1..641  
/organism="Trypanosoma brucei"  
/strain="TREU927/4 GUTat 10.1"  
/db\_xref="taxon:5691"  
/clone="Sheared DNA-46J23"  
/clone\_lib="Sheared DNA"





```

Qy 1730 attaatgaattaaataagatt 1750
|||||
Db 680 AATAATAATAATAATAATAAT 700

RESULT 3
BH401326
LOCUS
DEFINITION AG-ND-155A14,TF ND-TAM Anopheles gambiae genomic clone AG-ND-155A14
DNA linear GSS 11-DEC-2001
ACCESSION BH401326
VERSION BH401326
KEYWORDS 687 bp
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae
; Anopheles.
REFERENCE 1 (bases 1 to 687)
AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: AG-ND-155A14.TR
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjlloftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seg primer: M13 For
Class: BAC ends
Location/Qualifiers
source
1. .687
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-155A14"
/clone_lib="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT 339 a 62 c 69 g 217 t
ORIGIN

Query Match 2.0%; Score 92.8; DB 12; Length 687;
Best Local Similarity 51.4%; Pred. NO. 9.le-07;
Matches 214; Conservative 0; Mismatches 202; Indels 0; Gaps 0;

Qy 1390 aattattagtcacagatcacacagctgctcgacaaatcaagaatcatctgctac 1449
|||||
Db 28 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 87

Qy 1450 taacaatactgaaatgctgttgaatgaaagcaaaaataactaacaataactgaaatgc 1509
|||||
Db 88 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 147

Qy 1510 tgtgttgaatgaaacaaataactaacaatactgaaatgctgttgaatgaaacaa 1569
|||||
Db 148 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 207

Qy 1570 aaataactaacacagaaacgaataatagtcataatgaattgaactaataatgaacacc 1629
|||||
Db 208 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 267

Qy 1630 atcagccgctactcaagcaaaacttgaaagactaaatccctcaagctgctgaagctgttca 1689
|||||
Db 268 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 327

Qy 1690 aaatgccaaagattgtagccggttagtttaacagatgatcaaaatgaatgaataaataagat 1749
|||||
Db 328 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 387

Qy 1750 taactctctagctgctgaaaggggtgcaaaatgacaccttaagagacttagagg 1805
|||||
Db 388 TAATGGCTACAGCTTGTGATGAAGAGGGGTTTCATTGGTGACTCAGACACATGTAGG 443

RESULT 4
CNS04AEE
LOCUS
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
095G24 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL281759.1
VERSION AL281759.1
KEYWORDS GI:8020086
SOURCE GSS: genome survey sequence.
ORGANISM Tetraodon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 806)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 806)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bouneau,L., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 806)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
Location/Qualifiers
1. .806
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="095G24"
/clone_lib="G"
/note="Genoscope sequence ID : COBG095BD12LP1-end : T7"
BASE COUNT 247 a 87 c 99 g 333 t 40 others
ORIGIN

Query Match 2.0%; Score 91.4; DB 12; Length 806;
Best Local Similarity 43.9%; Pred. No. 1.6e-06;
Matches 240; Conservative 30; Mismatches 277; Indels 0; Gaps 0;

Qy 1388 aaaaattattagtgctaacagtaatacaaccagtggtcgacaaatcaagaatcatctgct 1447
|||||
Db 806 ACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 747

Qy 1448 actaacaatactgaaatgctgttgaatgaaagcaaaaataactaacaataactgaaaaa 1507
|||||
Db 746 AACAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 687

```



Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through ATCC. Sheared DNA end sequences search page: <http://www.tigr.org/tldb/mbd/tbdb/>.

Seq primer: M13-Forward

Class: Shotgun.

Location/Qualifiers

#### FEATURES

source

1. 729

/organism="Trypanosoma brucei"

/strain="TREU927/4 GUTat 10.1"

/db\_xref="taxon:5691"

/clone="Sheared DNA-49K6"

/clone\_lib="Sheared DNA"

/note="Vector: pUC18; Site\_1: SmaI; Constructed at The

Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from a cloned population of

Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically

sheared to give a tight size distribution (approx 2 kb).

The v + i method used for the library construction is

described in detail in Smith, H.O. and Venter, J.C.

(Making small insert libraries for whole genome shotgun

sequencing projects. In Genome Sequencing: A Practical

Approach, eds. M. Vaudin and B. Borell, Oxford University

Press, 1999).

360 a 58 c 58 g 253 t

BASE COUNT

ORIGIN

#### Query Match

Best Local Similarity 1.9%; Score 89.4; DB 12; Length 729;

Matches 198; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

QY 1359 aagctaattggtcaaatgctgtagtggaataattattagtgctaacagtaatacaacca 1418

Db 349 AAGTTATTGTGTATATAGCAGTAATAATAATAATAATAATAATAATAATAATAATA 408

QY 1419 gtggctgcacaaatcaagaatcatctgctactcaacaaatctgctggtgtaag 1478

Db 409 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 468

QY 1479 aaagcaaaataactaaactgctgctgctgctgctgctgctgctgctgctgctgct 1538

Db 469 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 528

QY 1539 atactgaaatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1598

Db 529 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 588

QY 1599 gtcaattaaagttaactaaatgaacacacacacacacacacacacacacacacac 1658

Db 589 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 648

QY 1659 agctaaatcctcaagctgctgctgctgctgctgctgctgctgctgctgctgctgct 1718

Db 649 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 708

QY 1719 cagatgacaaatgaatga 1737

Db 709 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 727

#### RESULT 7

AU076358

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 469)

AU076358 469 bp mRNA linear EST 27-JUL-1999

AU076358 Dictyostelium discoideum SS (H. Urushihara) Dictyostelium

discoideum cDNA clone SSA642, mRNA sequence.

AU076358

AU076358.1 GI:5607256

Dictyostelium discoideum.

Dictyostelium discoideum

Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.

#### AUTHORS

Urushihara, H.

Developmental CDNA in Dictyostelium discoideum (1999)

Unpublished (1999)

COMMENT

Contact: Hideko Urushihara

Institute of Biological Sciences

University of Tsukuba

3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan

Email: t402hu@sakura.cc.tsukuba.ac.jp

PROJECT = Dictyostelium discoideum cDNA project in Japan.

Location/Qualifiers

1. 469

/organism="Dictyostelium discoideum"

/strain="AX4"

/db\_xref="taxon:44689"

/clone="SSA642"

/clone\_lib="Dictyostelium discoideum SS (H. Urushihara)"

/dev\_stage="slug"

/dev\_stage="slug"

253 a 30 c 26 g 157 t 3 others

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 1.8%; Score 84.6; DB 9; Length 469;

Matches 177; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

QY 1429 aaatcaagaatcattgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1488

Db 5 AAATCAAGAATCAGTTTCTCCAGACATTATAATAATAATAATAATAATAATAATAATA 64

QY 1489 tactaaactactgaaaaatgctgtttaaataaataaataaataaataaataaataaata 1548

Db 65 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 124

QY 1549 tgcgtgtttaaataaataaataaataaataaataaataaataaataaataaataaata 1608

Db 125 TATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 184

QY 1609 gtaactaataatgaac 1668

Db 185 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 244

QY 1669 taaagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1728

Db 245 TAANAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 304

QY 1729 aattaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 1761

Db 305 TAATAATAATGTAATAATGTAATAATGTAATAATGTAATAATGTAATAATGTAATA 337

#### RESULT 8

AQ639568/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Trypanosoma brucei.

Trypanosoma brucei

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma.

1 (bases 1 to 401)

El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C.,

Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J.

, Fraser, C. and Adams, M.

Determination of clone end sequences from Trypanosoma brucei TREU

927/4 Fl library

Unpublished (1999)

Other\_GSSs: 927P1-4D3.TV

Contact: Najib M. El-Sayed

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: nelsayed@tigr.org  
 For clone/filter availability, please contact Sara Melville  
 (sml60@ole.bio.cam.ac.uk). Pl end sequences search page:  
 http://www.tigr.org/tdb/mdb/tbdb/.  
 Seq primer: SP6  
 Class: Pl ends

## FEATURES

source Location/Qualifiers  
 1. 401  
 /organism="Trypanosoma brucei"  
 /strain="TREU927/4"  
 /db\_xref="taxon:5691"  
 /clone="927Pl-4D3"  
 /clone\_lib="927Pl"  
 /note="Vector: pAD10SacBII; Site\_1: Bam HI; Constructed by Sara Melville, University of Cambridge, UK and Nancy Shepherd, Dupont Merck, Wilmington, DE. Genomic DNA was isolated from Trypanosoma brucei (stock TREU927/4) and partially digested with Sau 3AI. DNA fragments were cloned into the Bam HI site of pAD10SacBII vector (Genbank accession U09128). The average insert size is 65 Kb. Coverage: approx 4.4 X the haploid non-minichromosomal genome"  
 BASE COUNT 130 a 39 c 7 g 225 t  
 ORIGIN

Query Match 1.8%; Score 82.6; DB 12; Length 401;  
 Best Local Similarity 55.4%; Pred. No. 6.1e-05;  
 Matches 160; Conservative 0; Mismatches 129; Indels 0; Gaps 0;  
 QY 1345 taataactgaacgctgaagctgaatggtcgaattggtgtagatgggaataattattgtagctaa 1404  
 Db 395 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 336  
 QY 1405 cagtaatacaaccagtggtgcgacaatacaagaatcatctgctactaacaatactgaaaa 1464  
 Db 335 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 276  
 QY 1465 tgcgtgttgaatgaagcgaataactaacaatactggaatgctgtgttgaatgaaaa 1524  
 Db 275 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 216  
 QY 1525 caaaatactacaatactgaatgctgtgttgaatgaaaaaataactaacaacac 1584  
 Db 215 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 156  
 QY 1585 agaaacgataatagtcaattaaagttactaataatgaacacacatca 1633  
 Db 155 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 107

RESULT 9  
 LOCUS AQ941683 621 bp DNA linear GSS 27-JAN-2000  
 DEFINITION Sheared DNA-43D2-TR Sheared DNA Trypanosoma brucei genomic clone  
 Sheared DNA-43D2, DNA sequence.  
 ACCESSION AQ941683  
 VERSION AQ941683.1 GI:6764948  
 SOURCE GSS.  
 ORGANISM Trypanosoma brucei.  
 Trypanosoma brucei.  
 Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;  
 Trypanosoma.

REFERENCE 1 (bases 1 to 621)  
 AUTHORS El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C., Gerard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J., Fraser,C. and Adams,M.  
 TITLE Determination of clone end sequences from Trypanosoma brucei GUTat 10.1 sheared DNA library  
 JOURNAL Unpublished (1999)

## COMMENT

Other\_GSSs: Sheared DNA-43D2.TF  
 Contact: Najib M. El-Sayed  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: nelsayed@tigr.org  
 Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through ATCC. Sheared DNA end sequences search page: http://www.tigr.org/tdb/mdb/tbdb/.  
 Seq primer: M13-Reverse  
 Class: shotgun.

## FEATURES

source Location/Qualifiers  
 1. 621  
 /organism="Trypanosoma brucei"  
 /strain="TREU927/4 GUTat 10.1"  
 /db\_xref="taxon:5691"  
 /clone="Sheared DNA-43D2"  
 /clone\_lib="Sheared DNA"  
 /note="Vector: pUC18; Site\_1: SmaI; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (approx 2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds: M. Vaudin and B. Barrell, Oxford University Press, 1999)."  
 BASE COUNT 318 a 15 c 78 g 210 t  
 ORIGIN

Query Match 1.8%; Score 81.2; DB 12; Length 621;  
 Best Local Similarity 53.9%; Pred. No. 0.00011;  
 Matches 167; Conservative 0; Mismatches 143; Indels 0; Gaps 0;  
 QY 1314 aaagtaacaattacacaggtgtgctgaagttaactactgaacgtcaagtcgaatggcaca 1373  
 Db 299 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 358  
 QY 1374 ttgctgtgtagtggaaaaattattagtcgaacagtaatacaaccagtggtcgacaaatc 1433  
 Db 359 ATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 418  
 QY 1434 aagaatcatctgctactaacaatactgaaaatgctgtgttaataatgaacacaaatac 1493  
 Db 419 AAAGCAGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 478  
 QY 1494 acaatactgaaaatgctgtgttaataatgaacacaaatactactaatactgaaaatgc 1553  
 Db 479 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 538  
 QY 1554 ttgttaataatacaaaaaataactaacaacacagaaaaacgataataatagtcattaa 1613  
 Db 539 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 598  
 QY 1614 ctaataatga 1623  
 Db 599 ATAATAATAA 608  
 RESULT 10  
 LOCUS AU086536 500 bp mRNA linear EST 27-JAN-2001  
 DEFINITION AU086536 Sugano Malaria cDNA library Plasmodium falciparum CDNA clone XPf02783, mRNA sequence.  
 ACCESSION AU086536  
 VERSION AU086536.1 GI:12388677  
 KEYWORDS EST.

```

SOURCE      malaria parasite P. falciparum.
ORGANISM    Plasmodium falciparum
REFERENCE   1 (bases 1 to 500)
AUTHORS     Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.
TITLE       FULL-malaria: a database for a full-length enriched cDNA library
            from human malaria parasite, Plasmodium falciparum
JOURNAL     Nucleic Acids Res. 29 (1), 70-71 (2001)
MEDLINE     20574754
COMMENT     Contact: Junichi Watanabe
            Institute of Medical Science
            The University of Tokyo, Department of Parasitology
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Tel: 81-3-5449-5378
            Fax: 81-3-5449-5410
            Email: jwatanabe@manage.ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
            S. Construction and characterization of a full length-enriched and
            a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).
            Location/Qualifiers
FEATURES             1..500
source               /organism="Plasmodium falciparum"
                    /strain="3D7"
                    /db_xref="taxon:5833"
                    /clone="XPf2783"
                    /clone_lib="Sugano Malaria cDNA library"
                    /dev_stage="erythrocytic stage"
BASE COUNT          278 a 43 c 34 g 142 t 3 others
ORIGIN

Query Match      1.7%; Score 81; DB 9; Length 500;
Best Local Similarity 49.1%; Pred. No. 0.00012;
Matches 213; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

QY 1254 cattggtcagctcaattttaaaggagggttgtaacgcgtcgtgatcaagtag 1313
Db 57 CTTTATATATCTATATAGAAATGGAAATGATAATATATGCAATTTATCCGGTAGT 116
QY 1314 aaagtaacaattacaacggtgttgctgaagttaactactgaacgtaagtcgaatggtcaaa 1373
Db 117 CAAATATATAGAAATTAANAATAATGAAATAAATAATAATATGAAATAATAATAATA 176
QY 1374 ttggcgtagatggaataattattagctgaacagtaatacaacagtggtcgacaatac 1433
Db 177 ATACAAACAATAATAATAACAATAATAATAATAATAATAACAATAATAATAATAATA 236
QY 1434 aagaatcatctgctactaacaatactgaaatgctgtgttaataatgaaacaaataacta 1493
Db 237 ACAATATATATAGCAACAATAATAATAATAGCAACAATAATAATATAGTAACAACATA 296
QY 1494 acaactcgtgaaatgctgtgttaataatgaaacaaataactaacaatactgaaatgctg 1553
Db 297 ATAGTACAAACAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 356
QY 1554 ttgttaatgaaacaaataactaacaacacagaaacagataatagtcatttaaaagttaa 1613
Db 357 ATAGTACAAACAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 416
QY 1614 ctaaatgaaacaccatcgccgctactcaagcaaacactggaagactgaaacccctcaag 1673
Db 417 ATAATAATTACAATAATAGTAACAATAATAATAATAATAATCATCATGATACCTTAAT 476
QY 1674 ctgctaaggctgtt 1687
Db 477 ATGATAATACTATT 490

RESULT 11
BH391959
LOCUS
DEFINITION  BH391959 773 bp DNA linear GSS 11-DEC-2001
            AG-ND-180F9.TF ND-TAM Anopheles gambiae genomic clone AG-ND-180F9,
            DNA sequence.

malaria parasite P. falciparum.
Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 500)
Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.
FULL-malaria: a database for a full-length enriched cDNA library
from human malaria parasite, Plasmodium falciparum
Nucleic Acids Res. 29 (1), 70-71 (2001)
20574754
Contact: Junichi Watanabe
Institute of Medical Science
The University of Tokyo, Department of Parasitology
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Tel: 81-3-5449-5378
Fax: 81-3-5449-5410
Email: jwatanabe@manage.ims.u-tokyo.ac.jp
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
S. Construction and characterization of a full length-enriched and
a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).
Location/Qualifiers
FEATURES             1..500
source               /organism="Plasmodium falciparum"
                    /strain="3D7"
                    /db_xref="taxon:5833"
                    /clone="XPf2783"
                    /clone_lib="Sugano Malaria cDNA library"
                    /dev_stage="erythrocytic stage"
BASE COUNT          278 a 43 c 34 g 142 t 3 others
ORIGIN

Query Match      1.7%; Score 81; DB 9; Length 500;
Best Local Similarity 49.1%; Pred. No. 0.00012;
Matches 213; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

QY 1254 cattggtcagctcaattttaaaggagggttgtaacgcgtcgtgatcaagtag 1313
Db 57 CTTTATATATCTATATAGAAATGGAAATGATAATATATGCAATTTATCCGGTAGT 116
QY 1314 aaagtaacaattacaacggtgttgctgaagttaactactgaacgtaagtcgaatggtcaaa 1373
Db 117 CAAATATATAGAAATTAANAATAATGAAATAAATAATAATATGAAATAATAATAATA 176
QY 1374 ttggcgtagatggaataattattagctgaacagtaatacaacagtggtcgacaatac 1433
Db 177 ATACAAACAATAATAATAACAATAATAATAATAATAATAACAATAATAATAATAATA 236
QY 1434 aagaatcatctgctactaacaatactgaaatgctgtgttaataatgaaacaaataacta 1493
Db 237 ACAATATATATAGCAACAATAATAATAATAGCAACAATAATAATATAGTAACAACATA 296
QY 1494 acaactcgtgaaatgctgtgttaataatgaaacaaataactaacaatactgaaatgctg 1553
Db 297 ATAGTACAAACAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 356
QY 1554 ttgttaatgaaacaaataactaacaacacagaaacagataatagtcatttaaaagttaa 1613
Db 357 ATAGTACAAACAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 416
QY 1614 ctaaatgaaacaccatcgccgctactcaagcaaacactggaagactgaaacccctcaag 1673
Db 417 ATAATAATTACAATAATAGTAACAATAATAATAATAATAATCATCATGATACCTTAAT 476
QY 1674 ctgctaaggctgtt 1687
Db 477 ATGATAATACTATT 490

RESULT 12
AQ656286
LOCUS
DEFINITION  AQ656286 520 bp DNA linear GSS 23-JUN-1999
            Sheared DNA-27K23.TR Sheared DNA Trypanosoma brucei genomic clone
            Sheared DNA-27K23, DNA sequence.

```

ACCESSION A0656286  
 VERSION A0656286.1 GI:5164054  
 KEYWORDS GSS.  
 SOURCE Trypanosoma brucei.  
 ORGANISM Trypanosoma brucei.  
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
 Trypanosoma.  
 REFERENCE 1 (bases 1 to 520)  
 AUTHORS El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C.,  
 Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J.,  
 Fraser, C. and Adams, M.  
 TITLE Determination of clone end sequences from Trypanosoma brucei GUTat  
 10.1 sheared DNA library  
 JOURNAL Unpublished (1999)  
 COMMENT Other GSSs: Sheared DNA-27K23.TF  
 Contact: Najib M. El-Sayed  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: nelsayed@tigr.org  
 Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared  
 DNA library constructed at TIGR. Clones will be available for  
 distribution through ATCC. Sheared DNA end sequences search page:  
 http://www.tigr.org/tldb/mdb/tbdb/.  
 Seq primer: M13-Reverse  
 Class: shotgun.

FEATURES  
 source  
 Location/Qualifiers  
 1..520  
 /organism="Trypanosoma brucei"  
 /strain="TREU927/4 GUTat 10.1"  
 /db\_xref="taxon:5691"  
 /clone\_lib="Sheared DNA-27K23"  
 /clone\_lib="Sheared DNA"  
 /note="Vector: pUC18; Site 1: SmaI; Constructed at The  
 Institute for Genomic Research (TIGR), Rockville, MD.  
 Genomic DNA isolated from a cloned population of  
 Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically  
 sheared to give a tight size distribution (approx 2 kb).  
 The v + i method used for the library construction is  
 described in detail in Smith, H.O. and Venter, J.C.  
 (Making small insert libraries for whole genome shotgun  
 sequencing projects. In Genome Sequencing: A Practical  
 Approach, eds. M. Vaudin and B. Borell, Oxford University  
 Press, 1999)."  
 BASE COUNT 263 a 44 c 79 g 134 t  
 ORIGIN

Query Match 1.7%; Score 80.2; DB 12; Length 520;  
 Best Local Similarity 55.6%; Pred. No. 0.00016;  
 Matches 154; Conservative 0; Mismatches 123; Indels 0; Gaps 0;  
 QY 1363 taatggtgcaattggcgtagatggaataattattagtgtaacagtaatacaaccagtg 1422  
 Db 4 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 63  
 QY 1423 ctgcgaataatcaagaatcgtctactaataactgaaatcgtggtgtaataaga 1482  
 Db 64 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 123  
 QY 1483 caaaataactaataactgtaataactggtgtaataactgaaacaaataactaataac 1542  
 Db 124 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 183  
 QY 1543 tgaataatgctgtgttaataagaacaaataactaatacaacagaaacagataatgca 1602  
 Db 184 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 243  
 QY 1603 attaaagttaactaataataacacacacacacacacacacacacacacacacac 1639  
 Db 244 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 280

RESULT 13  
 CNS04KUS/c  
 LOCUS  
 DEFINITION

CNS04KUS 541 bp DNA linear GSS 21-MAY-2000  
 Tetraodon nigroviridis genome survey sequence T7 end of clone  
 117023 of library G from Tetraodon nigroviridis, genomic survey  
 sequence.

ACCESSION AL295309  
 VERSION AL295309.1 GI:8033889  
 KEYWORDS GSS: genome survey sequence.  
 SOURCE Tetraodon nigroviridis.  
 ORGANISM Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 Tetraodontidae; Tetraodon.  
 REFERENCE 1 (bases 1 to 541)  
 AUTHORS Roest-Crolius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,  
 Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and  
 Weissenbach, J.

TITLE Characterization and repeat analysis of the compact genome of the  
 freshwater pufferfish Tetraodon nigroviridis

JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 541)  
 AUTHORS Roest-Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,  
 Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,  
 Saurin, W. and Weissenbach, J.

TITLE Human gene number estimate provided by genome wide analysis using  
 Tetraodon nigroviridis DNA sequence

JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 541)  
 AUTHORS Genoscope.

TITLE Direct Submission  
 JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases

COMMENT This sequence is a single read and was generated as part of a large  
 scale clone-end sequencing project of the Tetraodon nigroviridis  
 genome. For more information, please take a look at  
 http://www.genoscope.cns.fr/Tetraodon.

FEATURES  
 source  
 Location/Qualifiers  
 1..541  
 /organism="Tetraodon nigroviridis"  
 /db\_xref="taxon:99883"  
 /clone\_lib="117023"  
 /clone\_lib="G"  
 /note="Genoscope sequence ID : C0BGL17AH12LPI-end : T7"

BASE COUNT 152 a 88 c 50 g 214 t 37 others  
 ORIGIN

Query Match 1.7%; Score 79.4; DB 12;  
 Best Local Similarity 48.7%; Pred. NO. 0.00023;  
 Matches 136; Conservative 26; Mismatches 117; Indels 0; Gaps 0;

QY 1345 taatactgaacgtcaagctaattggcgaattggcgaattggcgaattggcgaattggcga 1404  
 Db 531 TMAATAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 472

QY 1405 cagtaatacaaccagtggtcgacaaatcaagaatcctctgctactaataactgaa 1464  
 Db 471 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 412

QY 1465 tgcgtgttaataagaagcaaaataactaataactgaaatgctgtgttaataaga 1524  
 Db 411 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 352

QY 1525 caaaataactaataactgaaatgctgtgttaataagaagcaaaataactaataacac 1584  
 Db 351 TMAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 292

QY 1585 agaaacagataatagtcatttaactaataactgaaatgctgtgttaataaga 1623  
 Db 291 HMAHMRHAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 253

[illegible]



Search completed: September 22, 2002, 01:51:41  
Job time: 8381 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2002, 05:22:06 ; Search time 180.95 seconds  
(without alignments)  
6290.498 Million cell updates/sec

Title: US-09-995-587A-10  
Perfect score: 4634  
Sequence: 1 gttacaaagacaaatttt.....gttactgttcgaaagctt 4634

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	75	1.6	3305	3	US-09-068-043-1
2	69.4	1.5	10317	3	US-09-058-746-1
3	69.4	1.5	10317	4	US-09-438-142-1
4	64.4	1.4	740	4	US-09-451-117-1
5	61.2	1.3	1365	2	US-08-870-827-4
6	61.2	1.3	2408	2	US-08-870-827-5
7	59.8	1.3	4507	2	US-08-568-459A-3
8	59.8	1.3	4507	2	US-08-487-826B-3
9	57.2	1.2	3504	2	US-08-760-797A-2
10	57.2	1.2	3504	2	US-08-932-929B-2
11	56.6	1.2	8920	2	US-08-446-855A-1
12	56.6	1.2	8920	4	US-09-150-741-1
13	56.2	1.2	1631	3	US-09-118-319-1
14	55.4	1.2	3504	2	US-08-760-797A-4
15	55.4	1.2	3504	4	US-08-932-929B-4
16	54.4	1.2	9636	1	US-08-323-170B-1
17	54.4	1.2	9636	4	US-08-954-441-1
18	53.4	1.2	2237	4	US-08-914-999-7
19	52.4	1.1	1110	1	US-08-257-073-11
20	51.2	1.1	7218	1	US-08-232-463-14
21	51.2	1.1	19124	2	US-08-487-826B-13
22	50.4	1.1	1243	3	US-09-009-620-1
23	50.4	1.1	2223	3	US-08-257-073-4
24	50.4	1.1	4896	4	US-09-210-361-3
25	50.2	1.1	5116	1	US-08-038-682-1
26	50.2	1.1	5116	1	US-08-302-832-1
27	50.2	1.1	5116	2	US-08-530-198-1

28 50.2 1.1 5116 2 US-08-469-880-1 Sequence 1, Appli  
29 50.2 1.1 5116 2 US-08-728-470-1 Sequence 1, Appli  
30 50.2 1.1 5116 2 US-08-617-697-1 Sequence 1, Appli  
31 50.2 1.1 5116 4 US-08-719-641-1 Sequence 1, Appli  
32 50.2 1.1 9171 1 US-08-038-682-5 Sequence 5, Appli  
33 50.2 1.1 9171 1 US-08-302-832-5 Sequence 5, Appli  
34 50.2 1.1 9171 2 US-08-530-198-5 Sequence 5, Appli  
35 50.2 1.1 9171 2 US-08-469-880-5 Sequence 5, Appli  
36 50.2 1.1 9171 2 US-08-728-470-5 Sequence 5, Appli  
37 50.2 1.1 9171 2 US-08-617-697-5 Sequence 5, Appli  
38 50.2 1.1 9171 4 US-08-719-641-5 Sequence 5, Appli  
39 49.4 1.1 1956 4 US-08-559-896B-1 Sequence 1, Appli  
40 48.2 1.0 665 2 US-08-883-795A-36 Sequence 36, Appli  
c 41 48.2 1.0 5852 1 US-07-867-106-2 Sequence 2, Appli  
42 47.8 1.0 6744 1 US-08-119-125A-2 Sequence 2, Appli  
43 47 1.0 6124 4 US-08-213-419B-3 Sequence 3, Appli  
44 46.8 1.0 4248 3 US-08-678-614-1 Sequence 1, Appli  
45 46.6 1.0 1431 4 US-09-316-083-2 Sequence 2, Appli

## ALIGNMENTS

RESULT 1  
US-09-068-043-1  
; Sequence 1, Application US/09068043  
; Patent No. 6048694  
; GENERAL INFORMATION:  
; APPLICANT: MICHAEL GENE BRAMUCCI  
; APPLICANT: VASANTHA NAGARAJAN  
; TITLE OF INVENTION: POSITIVE SELECTION  
; TITLE OF INVENTION: VECTOR FOR BACILLUS SP.  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND  
; ADDRESSEE: COMPANY  
; STREET: 1007 MARKET STREET  
; CITY: WILMINGTON  
; STATE: DELAWARE  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 19898  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: DISKETTE, 3.5 INCH  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: MICROSOFT WINDOWS 95  
; SOFTWARE: MICROSOFT OFFICE 97  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/068,043  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/006,201  
; FILING DATE: NOVEMBER 3, 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FELTHAM, S. NEIL  
; REGISTRATION NUMBER: 36,506  
; REFERENCE/DOCKET NUMBER: CR-9807  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 302-992-6460  
; TELEFAX: 302-773-0164  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3305 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
US-09-068-043-1

Query Match 1.6%; Score 75; DB 3; Length 3305;  
Best Local Similarity 50.7%; Pred. No. 9.8e-09;





```

QY 2861 ataactcaggtgctgattaaactgcagtacctgcgaactggcgctactgtat 2920
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7627 AACAAACTGGCCTTGTGTAATAATGGGATCTTGATCCTAACGATGTAACCTTTACTTAC 7568
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2921 tctactatgcagtcactgtagc 2943
    ||||| ||||| ||||| |||||
Db 7567 TCACACTTCGCTGTACCTCAAGC 7545
    ||||| ||||| ||||| |||||

RESULT 3
US-09-438-142-1/c
; Sequence 1, Application US/09438142
; Patent No. 6258571
; GENERAL INFORMATION:
; APPLICANT: Ilya Chumakov
; APPLICANT: Hiroaki Tanaka
; TITLE OF INVENTION: High Throughput DNA sequencing vector
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/438,142
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: GENSET.015C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1017 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: synthetic DNA
; ORIGINAL SOURCE:
; ORGANISM: Cloning vector pGenDEL
; FEATURE:
; NAME/KEY: pGendel
; LOCATION: 1..10317
; FEATURE:
; NAME/KEY: Homology with X06404 compl (411..1668)
; LOCATION: 9..1266
; IDENTIFICATION METHOD: blastn against X06404
; FEATURE:
; NAME/KEY: Kanamycin resistance gene CDS
; LOCATION: 142..957
; IDENTIFICATION METHOD: By homology to X06404
; FEATURE:
; NAME/KEY: Tn1000's right end
; LOCATION: complement 1332..1371
; IDENTIFICATION METHOD: blastn against X60200)
; FEATURE:
; NAME/KEY: Homology with U46017 (1-472)
; LOCATION: 1423..1894
; IDENTIFICATION METHOD: blastn against U46017
; FEATURE:
; NAME/KEY: single stranded DNA replication origin
; LOCATION: 1423..1894
; IDENTIFICATION METHOD: By homology to U46017
; OTHER INFORMATION: mutation T -> C 1658

```

```

> NAME/KEY: CDS levensucrase gene
> LOCATION: complement 7379..8800
> IDENTIFICATION METHOD: By homology to X02730
> OTHER INFORMATION: mutation T -> C 7466
> OTHER INFORMATION: mutation A -> G 7739
> OTHER INFORMATION: mutation T -> C (Asn -> Asp) 8347
> OTHER INFORMATION: mutation T -> C 8600
> OTHER INFORMATION: mutation G -> A (Ala -> Val) 8772
> FEATURE:
> NAME/KEY: SLR3
> LOCATION: 8711..8731
> OTHER INFORMATION: Described in seqID 14
> FEATURE:
> NAME/KEY: Homology with J01636 (complement 1158..1465)
> LOCATION: 9298..9623
> IDENTIFICATION METHOD: blastn against J01636
> FEATURE:
> NAME/KEY: CDS alpha peptide beta-galactosidase
> LOCATION: complement 9276..9497
> IDENTIFICATION METHOD: By homology to J01636
> FEATURE:
> NAME/KEY: primer HE1
> LOCATION: complement 9465..9479
> FEATURE:
> NAME/KEY: primer HE2
> LOCATION: 9461..9475
> FEATURE:
> NAME/KEY: primer LacLRS2Avr
> LOCATION: complement 9603..9630
> FEATURE:
> NAME/KEY: primer LacE2M1u
> LOCATION: 9289..9314
> FEATURE:
> NAME/KEY: Homology with M77789 (1889..2576)
> LOCATION: 9629..10315
> IDENTIFICATION METHOD: blastn against M77789
> FEATURE:
> NAME/KEY: high copy-number double-stranded DNA replication origin
> LOCATION: complement 9629..10315
> IDENTIFICATION METHOD: By homology to M77789
> OTHER INFORMATION: mutation C -> T 9803
> OTHER INFORMATION: site ScaI 10029 - 10034
> OTHER INFORMATION: site PmlI 10038 - 10043
> OTHER INFORMATION: CLONING SITES 10031 - 10041
> FEATURE:
> NAME/KEY: oriLRd
> LOCATION: 9856..9881
> OTHER INFORMATION: Described in seqID 8
> FEATURE:
> NAME/KEY: OS1
> LOCATION: 10009..10026
> OTHER INFORMATION: Described in seqID 10
> FEATURE:
> NAME/KEY: OR1
> LOCATION: complement 10046..10062
> OTHER INFORMATION: Described in seqID 11
> FEATURE:
> NAME/KEY: oriLRr
> LOCATION: complement 10182..10202
> OTHER INFORMATION: Described in seqID 9
> US-09-438-142-1
>
> Query Match 1.58; Score 69.4; DB 4; Length 10317;
> Best Local Similarity 47.4%; Pred. No. 3.4e-07;
> Matches 381; Conservative 0; Mismatches 392; Indels 30; Gaps 5;
>
> Qy 2144 caagaatggtcaggttcagctattgtaaaatgatgatgataatgaatcatttttcaac 2203
> |||||
> Db 8320 CAAGAATGGTCAGGTTTCAGCCACATTTACATCTGACCGAAAAATCCGTTTATTCTACAC- 8260
> |||||
> Oy 2204 tcaaatgatgcgtctgattacaagtgtgaatgatcaaacgccttgttacgcgaacattaac 2263
> |||||

```

	Query Match	1.3%;	Score 61.2;	DB 2;	Length 1365;
	Best Local Similarity	50.0%;	Pred. NO. 1.5e-05;		
	Matches 181;	Conservative	0;	Mismatches 178;	Indels 3; Gaps 1;
QY	2299	tgttgataattacaagtttggttgaagtgatggatttcactaccacaaacttgaaca	2358		
Db	591	TGATCATAAATCTGTCTTTGATGGCGAGAGCGGTACAGTTTATCAAAATATTACGCAATT	650		
QY	2359	attcgaaacggcaaaagatggtgaaaaatgatgattactgcttcagtgaccacacagttgt	2418		
Db	651	TATCGATGAAGGCAAGTGGATTTCAGGTGATACCACTATTAGACAGCCCTCACTATGT	710		
QY	2419	tcaattgaaaaatggtgatcggttatctctgttatbtaagagctaatactgggacagaagatta	2478		
Db	711	TGA---AGATGAAGGCCATAAATATCTGTCTTTGAAGCGAATACTGGAACAACAGATGG	767		
QY	2479	ccaaagtgcagcaccaaatttataattgggctaaactatggcgagatgatcccttcaatat	2538		
Db	768	TTATCAAGGCGATCAGTCTTTCATATAAAGCTTACTATGGCGGAAGTGCAGCTTCTT	827		
QY	2539	taagagttccttcaagcttttgtaataataagaagagatcgctgaattggctggttttagctaa	2598		
Db	828	CCAGAATTGAAAAAATAAATCTGCTTCAAGTCTTAAAAACAATAATGCTCTCTTTAGCGAA	887		
QY	2599	tgtgcacttggatctttaagctcactaacaatcaagaatgaagccaaagttgaagaagt	2658		
Db	888	TGGTGCAATTAGGCATTGTTGAATTGGCCGATGACTATACAGTGAAGAGTGTATTGAACCC	947		
QY	2659	at 2660			
Db	948	AT 949			

```

RESULT      6
US-08-870-827-5
; Sequence 5, Application US/08870827
; Patent No. 5962297
; GENERAL INFORMATION:
; APPLICANT: Tsusaki et al.
; TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
; TITLE OF INVENTION: ACTIVITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,827
; FILING DATE: 06-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 170,630/1996
; FILING DATE: 10-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TSUSAKI=2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2408 base pairs
; TYPE: nucleic acid

```

STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: Bacillus sp.  
INDIVIDUAL ISOLATE: V230 (FERM BP-5054)  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 1..360  
IDENTIFICATION METHOD: E  
NAME/KEY: signal peptide  
LOCATION: 361..456  
IDENTIFICATION METHOD: S  
NAME/KEY: mat peptide  
LOCATION: 457..1821  
IDENTIFICATION METHOD: S  
NAME/KEY: 3'UTR  
LOCATION: 1822..2408  
IDENTIFICATION METHOD: E  
US-08-870-827-5

Query Match 1.3%; Score 61.2; DB 2; Length 2408;  
Best Local Similarity 50.0%; Pred. No. 1.8e-05;  
Matches 181; Conservative 0; Mismatches 178; Indels 3; Gaps 1;

QY 2299 tgttgataattcaagttttgttggaagtgatgatttcactaccacaaacttatgaaca 2358  
Db 1047 TGATCATTAATCTGCTTTGATGGCGGAGACGTACAGTTTATCAAAATATTCAGCAATT 1106  
QY 2359 attcgcaacgcaagatcgtagaaatgatattactcttactgacgtgacccacacgttgt 2418  
Db 1107 TATCGATGAGGCAACTGATTCAGGTGATACCATCTTTTAAGAGACCCCTCACTATGT 1166  
QY 2419 tcaattagaaaatggtgactgtttatcttgcgaagtaatactggagcagaagattta 2478  
Db 1167 TGA---AGATAAGGGCCATAAATATCTTCTTTGAAGCGAATACTGGAACACACATGG 1223  
QY 2479 ccaagtgacgacccaatttataattgggctaactatggtggcgatgatgcctcaatat 2538  
Db 1224 TTATCAGCGGATGCTCTTTCAATAAATAAGCTTACTATGCGGAAGTGACGTCTTCCT 1283  
QY 2539 taagagttcttcaagcttttgaataataagaagatcgatgattggctgggttttagctaa 2598  
Db 1284 CCAGAAATGAAAAATAAACTGCTTCAAGTCTTAAAAACAAATGCTTCTTTAGCGAA 1343  
QY 2599 tgggtcacttggtatcttaaaagctcaactaacaatacaagtaagccaaaggttgaagaagt 2658  
Db 1344 TGGTGCAATTAGGCATTTGTAATTGGCCGATGACTATACAGTGAAAAAGTGTATGAAACC 1403  
QY 2659 at 2660  
Db 1404 AT 1405

RESULT 7  
US-08-568-459A-3  
; Sequence 3, Application US/08568459A  
; Patent No. 5849306  
; GENERAL INFORMATION:  
; APPLICANT: Sim, Kim L.  
; APPLICANT: Chitnis, Chetan  
; APPLICANT: Miller, Louis H.  
; APPLICANT: Peterson, David S.  
; APPLICANT: Su, Xin-zhaun  
; APPLICANT: Wellens, Thomas E.  
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
; NUMBER OF INVENTION: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe Martens Olson & Bear  
; STREET: 620 Newport Center Drive 16th Floor  
; CITY: Newport Beach

STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/568,459A  
FILING DATE: 07-DEC-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelson, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121.001CPI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4507 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium falciparum  
US-08-568-459A-3

Query Match 1.3%; Score 59.8; DB 2; Length 4507;  
Best Local Similarity 46.0%; Pred. No. 5e-05;  
Matches 202; Conservative 0; Mismatches 237; Indels 0; Gaps 0;  
QY 1308 aagtagaaagtaacaaatcaacacgggtgttgctgaagcttaatactactgaacgtlcaagctaaag 1367  
Db 2670 AAGATCAAAATAGTGATGAATCGGAAGAACTGTACTAAATCATATATCAAAAGTCCAT 2729  
QY 1368 gtcaaatggcgtagatgaaataattattagtgctaacagtaatacaacacagtggtctga 1427  
Db 2730 CTATAAATAATGGAGATGATTCAGGCAGTGGAAAGTCAACAGTGAATCTAGTAGTT 2789  
QY 1428 caaatcaagaatcaatctgctactaacaataactgaaataactgtgttgtaataagcaaaaa 1487  
Db 2790 CAATAGTACTGATTGCTATTGATGATGATGAGAAATGCTGATACATTTGTTTCGAACAAG 2849  
QY 1488 atactaacaatactgaaataactgtgttgtaataagaaacaaaaataactaactactgaaa 1547  
Db 2850 ATACAGCAAAATACTGAAGATGTTATTAGAAAAAATACTGTCACAGAGGATGAAGATGAAA 2909  
QY 1548 atgctgtgttgaataaacaacaaataactaacaacagaaacagataatagtcaattaa 1607  
Db 2910 AAGGCGAGATGAAGAAGACACATAGTACTTCTGAAAGCTTAAGTTCACTGAAGAAAAA 2969  
QY 1608 agttaactaataatgaacaacccatcagccgctactcaagcaaaacttgagaagctaaatc 1667  
Db 2970 TGTAACTGATAATGAAGGAGGAAATAGTTTAAATCATGAAGAGGTGAAGACATCTACTA 3029  
QY 1668 ctcaagctgctaaagctgttcaaaaatgccaaagattgatcgccggtgtttaaacaagatgc 1727  
Db 3030 GTAATTCGTAAATGTTCAACAGTCTGGAGGAATTTGTTAATATGATGTTGGAAGAAGAC 3089  
QY 1728 aaattaatgaattaaataa 1746  
Db 3090 TAAAGATACTTTAGAAAA 3108  
RESULT 8  
US-08-487-826B-3  
; Sequence 3, Application US/08487826B  
; Patent No. 5993827



GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellem, Thomas F.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelsen, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121.001CPI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4507 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium falciparum  
US-08-487-826B-3

[illegible]

```

Qy 1668 ctcaagctgctaaagcgttccaataatgccaaatgagtcoggttagtttaacaagatgac 1727
Db 3030 GTAATTCGTGAATATGTTCAACACGCTCTGGAGGAATTTGTTAATATGAATGTTGAGAAAGAAC 3089
Qy 1728 aaattaatgaattaataa 1746
Db 3090 TAAAGATACCTTTAGAAAA 3108

RESULT 9
US-08-760-797A-2
; Sequence 2, Application US/08760797A
; Patent No. 5928902
; GENERAL INFORMATION:
; APPLICANT: De Wilde, Michel
; APPLICANT: Cohen, Joseph
; TITLE OF INVENTION: Hybrid Protein Between CS
; TITLE OF INVENTION: from Plasmodium and HBSAG
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/760.797A
; FILING DATE: 04-DEC-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/442,612
; FILING DATE: 17-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Baumeister, Kirk
; REGISTRATION NUMBER: 33,833
; REFERENCE/DOCKET NUMBER: B45015-1C2
; TELEPHONE: 610-270-5096
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3504 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-760-797A-2

```

Query Match	1.2%	Score 57.2:	DB 2:	Length 3504;
Best Local Similarity	47.7%	Pred. No. 0.00019;		
Matches 207;	Conservative	0;	Mismatches 218;	Indels 9; Gaps 1;
QY 1462	aaatgctgtgttaatgaaagcaaaaatactaaccaatactgaaatgctgtgttaatga	1521		
Db	1100	AAATGCAAAACCCAAACGCAAAACCCCAATGCAAAATCTTAATGCAAAACCCCAATGCAAAATCC	1159	
QY 1522	aaacaaaaatactacaatactgaaatgctgtgttaatgaaacaaaaatactaacaa	1581		
Db	1160	TAATGCAAAATCTTAATGCAAAATCCAAATGCAAAATCTCAAAATGCAAAACCCAAACGCAAAACCC	1219	
QY 1582	caagaaaaacgataagttaacattaaagttaactaaataatgaaacacacatcagcgctac	1641		
Db	1220	CAATGCAAAATCTTAATGCCCAATCCCAATGCAAAATCTCAAAATGCAAAACCCCAATGCAAAACCC	1279	
QY 1642	tcaagcaaaacttgaag-----aagctaaatctcgaagctgctaaggctgttcaaaa	1692		



US-08-446-855A-1

Query Match 1.2%; Score 56.6; DB 2; Length 8920;  
Best Local Similarity 46.4%; Pred. No. 0.00039;  
Matches 185; Conservative 0; Mismatches 214; Indels 0

Qy	1429	aattcaagaatcatctgctaactaacataactgaaaagtgttgttaattaagaaacaaaa	1488
		- - - - -           - - - - -	
Db	6247	AAATAAFAAAAAATTAAGNAAAAAAAAAATTGCAAAAANAANTAACGGGAANAANTAACACAANAA	6306

Qy	1489	tactaacaaactgaaatgctgtttaaataagaacaaataactaatactgaaaa	1548
Db	6307	GAATTATCATTTATGTTTAATCATAAAGGAATAATGAATTAATAGTAACAAATATTGAATC	6366

[illegible][illegible]

**QY** 1669 tcaagctgctaaggctgttcaaagtccaagaattgatgcggtagtttaacagatgatca 1728  
| | | | | | | | | | | | | | | |  
**Db** 6487 TTTAGTACTGAAGAATTGAACTAGTGATGATGATGATTAATTATTCATCATCAAGA 6546  
| | | | | | | | | | | | | | | |

**QY** 1729 aattaatgaattaaataagattacttcttaagtctgtgaaaagggtgcacaaattgac 1788  
| | | | |  
**Dδ** 6547 TGAAGAAGTACACTATGACCGATGATAATGTGATGATCATGCATCATCATCATGCCCA 6606

QY 1789 ctttaaggaccttagaggggattgggaatgcattgttaa 1827  
          |||||               ||          |||||          |  
Db 6607 GGATGAGGAGGATAATCATTTATTATATATGATGATGGTTA 6645

RESULT 12  
US-09-150-741-1

Sequence 1, Application US/09150741  
; Patent No. 6183996  
; GENERAL INFORMATION:  
; APPLICANT: Stewart et al.

```

; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
; Patent No. 6183996
; TITLE OF INVENTION: Synthetase II
; FILE REFERENCE:

```

```

; CURRENT APPLICATION NUMBER: US/09/150,741
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PL6380
; EARLIER FILING DATE: 1992-12-16

```

```

; EARLIER APPLICATION NUMBER: AU93/00617
;
; EARLIER FILING DATE: 1993-12-02
;
; EARLIER APPLICATION NUMBER: 08/446,855
;
; EARLIER FILING DATE: 1995-07-06
;

```

```

; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 8920

```

; TYPE: DNA  
; ORGANISM: Plasmodium falciparum  
US-09-150-741-1

Query Match 1.2%; Score 56.6; DB 4; Length 8920;  
Best Local Similarity 46.4%; Pred. No. 0.00039;  
Matches 185; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

DQ 1429 aaatcaagaatcatctgctactacaatactgaaaatgctgttgaatgaagcaaaaa 1488  
||||| |||| | | | | | | | | | | | | | |  
Db 6247 aaattaaaanaaaaataaagaagaaaaaatcggaagaaaaatacgaacccccccccc  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1489 tactaacaatactgaaaaatgctgtgttgaataagaacaaaaataactaacaatactgaaa 1548

[illegible]

## RESULT 13

```

US-09-118-319-1
; Sequence 1, Application US/09118319
; Patent No. 6114158
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Chen, Huizhong
; APPLICANT: Ljungdahl, Lars G.
; TITLE OF INVENTION: Orpinomyces Cellulase Celf Protein and Coding Sequences
; FILE REFERENCE: 33-98sequence listing
; CURRENT APPLICATION NUMBER: US/09/118.319
; CURRENT FILING DATE: 1998-07-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1631
; TYPE: DNA
; ORGANISM: Orpinomyces sp. PC-2
; FEATURE:
; NAME/KEY: intron
; LOCATION: (187)..(297)
US-09-118-319-1

```

Query Match	1.2%	Score 56.2;	DB 3;	Length 1631;
Best Local Similarity	48.0%	Pred. No. 0.00025		
Matches 160;	Conservative	0;	Mismatches 173;	Indels 0;
				Gaps 0

QY	1439	tcactgtctactaacaatactgaaatgctgtt	taataagacacaaaataactaacat	1498	
Db	386	tcactcaaatccactcttaataa	gctctcaacaataataacatgataacacaac	445	
QY	1499	actgaaaaatgctgtt	gttaagaaaaacaaaataactaacaatactgaaaaatgctgtt	1558	
Db	446	acaacacacaataataataata	taataataataataataataactctgtt	505	
QY	1559	aatgaaaaacaaaatact	aacaacacagaaacgataatagtc	aattaaagttaactaat	1618
Db	506	agtggaagtactaaaacttct	caactaatcaaatattgct	taatccaaaattcaatgaa	565
QY	1619	aatgacaacacatacgcg	ctactcaagcaactggaagagctaaatctccaagctgt	1678	
Db	566	gaagtcaattctcttatt	ccaaaataagttgggacttcacacaaaaggtccaaaagtt	625	
QY	1679	aaggctgttcaaaatgcc	agattgatgccggtagtttaacagatgatcaaat	taatgaa	1738
Db	626	aaggatgttccaactgc	tgttggttagcttgggaaggtgtcc	aggtggaagttgaacaa	685
QY	1739	ttaataagaattaaacttct	ctctaagctctgctgaa	1771	



Qy	1563	aaacacaaaatactaacacacacagaaaacgataatagtc	aaattaaagttaactaactaataatg	1622
Db	1189	CAAAACCCAAACGCAAAACCCCAATGCAATCTCTAATGCCAATCCAAATGCAAAATCCCAATG		1248
Qy	1623	aaacaccatacagcgctactcaagcaaaacttgaag-----aa	gctaaaactctctcaag	1673
Db	1249	CAAAACCCCAATGCAAAACCCCAATGCAAAACCCCAATGCAAAATCTCTAATATAAAACAATCAAG		1308
Qy	1674	ctgctaagcgctgtctaaaaatgccaaagattgatgcgcggtagtttaac	agatgatcaaaatta	1733
Db	1309	GTAAATGGACAAAGTGCACAATATGCCAAATGACCCAAACCGAAATGTAGATGAAATGCTA		1368
Qy	1734	atgaattaaataagattaacttctctaagctgcgtgaaaagggtgc	aaaatttgaccttta	1793
Db	1369	ATGCCAACAGTGTGTAAAAAATATATATACGAAGAACCAGTGTGTAAGCACATAAAG		1428
Qy	1794	aggacttagagggtatgtgtaatgcctattgttaagcaagatcc	acacataatgctattcctt	1853
Db	1429	AATATTATAACAAAATACAAAATCTCTTTCAACTGAATGGTGTCCCATGTAGTGTAACTT		1488
Qy	1854	attctaatgctaaaggaatcaagaatatgcctg		1886
Db	1489	GTGGAATGGTATTCAAGTTAGAAATTAAGCCGTG		1521

Search completed: September 22, 2002, 05:23:26  
Job time: 20741 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2002, 09:38:17 ; Search time 1031.3 Seconds  
(without alignments)  
7714.711 Million cell updates/sec

Title: US-09-995-587A-10

Perfect score: 4634

Sequence: 1 gttacaagaacaaaatttt.....gttactgcttgcacaaagctt 4634

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_032802.\*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88.6	1.9	2079	23 AAS85786	DNA encoding novel
c 2	88.6	1.9	2079	23 AAS85998	DNA encoding novel
3	88.6	1.9	2079	23 AAS87342	DNA encoding novel
c 4	88.6	1.9	2079	23 AAS88703	DNA encoding novel
5	80.8	1.7	5940	21 AAA70105	Plasmodium falciparum
6	78.2	1.7	3947	23 AAS86050	DNA encoding novel
c 7	77.2	1.7	6642	23 AAS81427	DNA encoding novel
c 8	76.6	1.7	979	23 AAS86024	DNA encoding novel
c 9	76	1.6	5340	22 AAF26289	P. falciparum eg5

10	75.6	1.6	3990	23 AAS89967	DNA encoding novel
11	75.6	1.6	3990	23 AAS92589	DNA encoding novel
12	75.6	1.6	6033	23 AAS85839	DNA encoding novel
13	75.4	1.6	1866	23 AAS86014	DNA encoding novel
14	75.4	1.6	1866	23 AAS87349	DNA encoding novel
15	75	1.6	1032	23 AAS86020	DNA encoding novel
16	75	1.6	1099	23 AAS74868	DNA encoding novel
17	75	1.6	1188	23 AAS68479	DNA encoding novel
18	75	1.6	1188	23 AAS74483	DNA encoding novel
19	75	1.6	1188	23 AAS6028	DNA encoding novel
20	75	1.6	1188	23 AAS87344	DNA encoding novel
c 21	75	1.6	1212	23 AAS86040	DNA encoding novel
22	75	1.6	1245	23 AAS86037	DNA encoding novel
23	75	1.6	1281	23 AAS68475	DNA encoding novel
24	75	1.6	1281	23 AAS87378	DNA encoding novel
25	75	1.6	1284	23 AAS65782	DNA encoding novel
26	75	1.6	1284	23 AAS74491	DNA encoding novel
27	75	1.6	1453	23 AAS69429	DNA encoding novel
28	75	1.6	1459	23 AAS85820	DNA encoding novel
c 29	75	1.6	1512	23 AAS86017	DNA encoding novel
30	75	1.6	1512	23 AAS88769	DNA encoding novel
c 31	75	1.6	1535	23 AAS86038	DNA encoding novel
c 32	75	1.6	1551	23 AAS86140	DNA encoding novel
33	75	1.6	1563	23 AAS84054	DNA encoding novel
34	75	1.6	1584	23 AAS68494	DNA encoding novel
35	75	1.6	1629	23 AAS85800	DNA encoding novel
c 36	75	1.6	1629	23 AAS86022	DNA encoding novel
c 37	75	1.6	1629	23 AAS87359	DNA encoding novel
38	75	1.6	1674	23 AAS86041	DNA encoding novel
39	75	1.6	1690	23 AAS85792	DNA encoding novel
40	75	1.6	1690	23 AAS86019	DNA encoding novel
41	75	1.6	1690	23 AAS87350	DNA encoding novel
42	75	1.6	1810	23 AAS85801	DNA encoding novel
43	75	1.6	1810	23 AAS86023	DNA encoding novel
44	75	1.6	1827	23 AAS84506	DNA encoding novel
45	75	1.6	1827	23 AAS85069	DNA encoding novel

#### ALIGNMENTS

RESULT 1  
AAS85786  
ID AAS85786 standard; cDNA; 2079 BP.

XX AAS85786;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #21590.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG21599.

PT New isolated polynucleotide and encoded polypeptides, useful in



Query Match 1.9%; Score 88.6; DB 23; Length 2079;  
Best Local Similarity 48.9%; Pred. NO. 5.9e-09;  
Matches 331; Conservative 0; Mismatches 334; Indels 12; Gaps 3;

QY 1976 gttatcaattacatggtttatcaactagttattgctatgaggggtattccaaattcgcca 2035  
Db 1827 GTAGTGAATACACGGGTATACAGCTTGTGTTGCTTCTTGGCGGAACCCGAAAGACGCTG 1768  
QY 2036 actggagataatcatatctcttttacaacaagatggtgataagacttttctcat 2095  
Db 1767 ATGACACATCACTACATGTTTATCAAAAGCTGGAAGAACGCGGCGTCTTTAAA 1708  
QY 2096 tggcgcaatgcaggttcaaatcttgggaactaaagaacaaatggtgttccagaagtgtca 2155  
Db 1707 GACAGCGATAAGTTCGACGCCAACGATCCGATCCTGAAAGATCAGACGCAAGAAATGGTCC 1648  
QY 2156 ggttcagctattgtaaatgatggttacaattcaactattttcaacctcaaatgatacg 2215  
Db 1647 GGTTCGCAACCTTTACATCTGACGGAAATCCGTTTATTCTACAC-----TGACTAT 1594  
QY 2216 tctgattacaagttgaatgatcaacgccttgcaccgcaacattaaaccttaattgtgat 2275  
Db 1593 TCCGGTAAACATTACGGCAACAAGCTGACACACGCGCAGGTAATGTGTCAAAATCT 1534  
QY 2276 gataacggtgttcaatcaagagtggtgataattatcaagtttgggttgaagggtgagga 2335  
Db 1533 GATGACACACTCAAAATCAACGGAGTGAAGATCACAAAACGATTTTTCGCGGAGACGGA 1474  
QY 2336 ttctactaccaacttatgaacattcgcaacgcaacgcaagatcgtaaaatgatgattac 2395  
Db 1473 AAACATATCAGAAGCTTACAGCTTTATCGATGAAGCAATTTATACATCGCGGACAAAC 1414  
QY 2396 tgcctacgtgaccacacgctgtttcaatgaataatggtgatcgttcttctgtattcgaa 2455  
Db 1413 CATACGCTGAGAGACCTCACTACGTTGAGACAAAGGCCATAAATACCTTGTATTGAA 1354  
QY 2456 gctaaactgggacagaa---gattacaaagtgcacacacaaatataatggtggctaac 2512  
Db 1353 GCCAACACGGGAACAGAAACGGATACCAAGCGGCAAGAAATCTTTATTAAACAAAGCGTAC 1294  
QY 2513 tatggtggcgatgagtccttcaatataaagagttccttcaagcttttgaataataaagaag 2572  
Db 1293 TACGGCGGCGCAGCAACTTCTCCGTAAGAAAGCCAGCAAGCTTCACGACGCGCTAAA 1234  
QY 2573 gategtaattggctgttagttagtaattggtgactgtgattctttaaagctcaactaacaat 2632  
Db 1233 AAACGCGA---TGCTGAGTTAGCGAAGCGGCCCTCGGTATCATAGAGTTTAAATAATGAT 1177  
QY 2633 caaagtaagcacaaggt 2649  
Db 1176 TACACATTGAAAAAGT 1160

RESULT 3  
ID AAS87342  
XX AAS87342 standard; CDNA; 2079 BP.  
AC AAS87342;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #23146.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN W0200175067-A2.  
XX  
PD 11-OCT-2001.  
XX

30-MAR-2001; 2001WO-US08631.  
31-MAR-2000; 2000US-0540217.  
23-AUG-2000; 2000US-0649167.  
(HYSE-) HYSEQ INC.  
Drmanac RT, Liu C, Tang YT;  
WPI: 2001-639362/73.  
P-PSDB; ABG23155.  
New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity  
Claim 1; SEQ ID No 23146; 103pp; English.  
The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.  
Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
Sequence 2079 BP; 710 A; 433 C; 475 G; 461 T; 0 other;

Query Match 1.9%; Score 88.6; DB 23; Length 2079;  
Best Local Similarity 48.9%; Pred. NO. 5.9e-09;  
Matches 331; Conservative 0; Mismatches 334; Indels 12; Gaps 3;

QY 1976 gttatcaattacatggtttatcaactagttattgctatgaggggtattccaaattcgcca 2035  
Db 253 gtagtgaatacaacggctacgctgtgttcttgcgggaacccggaagacgctg 312  
QY 2036 actggagataatcatatctcttttacaacaagatggtgataaactttctcat 2095  
Db 313 atgacacataatctacatgttttataaaagctggaaacgcggcgctgtctttaa 372  
QY 2096 tggcgcaatgcaggttcaaatcttgggaactaaagaacaaatggttccagaagtgtca 2155  
Db 373 gacagcgataagttcgacgccaacgacccgctcctgaaagatcagacgcaagaatgctcc 432  
QY 2156 ggttcagctattgtaaatgatggtgataaactattttcaacctcaaatgatacg 2215  
Db 433 ggttgcacactttacatctgacgggaaaaatccgtttattctacac-----tgactat 486  
QY 2216 tctgattacaagttgaatgatcaacgccttgcaccgcaacattaaaccttaattgtgat 2275  
Db 487 tccggttaaacattacgcaacaaagcctgcacacagcgcaaggttaaatgtgtcaaaatct 546  
QY 2276 gataacggtgttcaatcaagagtggtgataattatcatcaagtttgggtgaggtgagga 2335  
Db 547 gatgacacactcaaaatcaacggagtggaagatcacaaaaagatttttgcgagacgga 606  
QY 2336 ttctactaccaacttatgaacaattcgcaacgcaagatcgtaaaatgatgattac 2395



Db 607 aaacatcatcagacgttccagcagtttatcatgatgaaggaatatacatcatccgagcaac 666  
 Qy 2396 tgcctacgtgaccacacagcgttgcataatagaataatggtgatcttctgtatttcgaa 2455  
 Db 667 catcacgtgagagaccctcactacgttgtagacacaaaggccataataacctgtatttcgaa 726  
 Qy 2456 gctaataactgggacagaa---gattaccaaagtacacacaaatttataatgggtaac 2512  
 Db 727 gcaacacggaacagaaacaggaatcaccaggaaggaatcttatttatacaaacgctac 786  
 Qy 2513 tatgggtggcagatgatcccttcaataatgaagagttccttcaagccttttgaataataagaag 2572  
 Db 787 tacggcgccgacgaacttctccgtataaagaacgaagcttcagcagcgctaaa 846  
 Qy 2573 gatcgtgaattggcgtggttagcctaagtgtgcaactggtgatcttataaagctcactaacaat 2632  
 Db 847 aaacgcga---tgctgagtgtagcgaacgagcgcctcgtgatcatagagttataataatgat 903  
 Qy 2633 caaagtaagccaaaggt 2649  
 Db 904 tacacattgaaaaaagt 920

## RESULT 4

AAS88703/C

ID AAS88703 standard; cDNA; 2079 BP.

AC AAS88703;

XX

DT 13-FEB-2002 (first entry)

DE

XX DNA encoding novel human diagnostic protein #24507.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

XX 30-MAR-2001; 2001WO-U508631.

XX

PR 31-MAR-2000; 2000US-0540217.

XX

XX 23-AUG-2000; 2000US-0649167.

XX

XX (HYSE-) HYSEQ INC.

XX

XX Drmanac RT, Liu C, Tang YT;

XX

XX WPI; 2001-639362/73.

XX

XX P-FSDB; ABG24516.

XX

New isolated polynucleotide and encoded polypeptides, useful in  
 diagnostics, forensics, gene mapping, identification of mutations  
 responsible for genetic disorders or other traits and to assess  
 biodiversity.

XX

XX Claim 1; SEQ ID No 24507; 103pp; English.

XX

The invention relates to isolated polynucleotide (I) and  
 polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 and gene mapping, and in recombinant production of (II). The  
 polynucleotides are also used in diagnostics as expressed sequence tags  
 for identifying expressed genes. (I) is useful in gene therapy techniques  
 to restore normal activity of (II) or to treat disease states involving  
 (II). (II) is useful for generating antibodies against it, detecting or  
 quantitating a polypeptide in tissue, as molecular weight markers and as  
 a food supplement. (II) and its binding partners are useful in medical  
 imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations in  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 2079 BP; 461 A; 475 C; 433 G; 710 T; 0 other;

Query Match 1.9%; Score 88.6; DB 23; Length 2079;  
 Best Local Similarity 48.9%; Pred. No. 5.9e-09;  
 Matches 331; Conservative 0; Mismatches 334; Indels 12; Gaps 3;

QY 1976 gtatcattacatgggttatcaactagttattgtctatgtgggtattccaaattccgcca 2035  
 Db 1827 GTAGCTGAATACAAACGGCTATCAGCTTGTGTTCTGCGGAGACCCGAAAGACGCTG 1768

QY 2036 actggagataatcatatctatctctttacaacaagtatggtgataatgacttttctcat 2095  
 Db 1767 ATGACACATCAATCTACATGTTTATCAAAAGCTGGAACACGGGCGCGTGTAAAA 1708

QY 2096 tggcgcaatgcaggttcaatcttttgaactaaagaacacaaatgtgtcccaagaatgtca 2155  
 Db 1707 GACAGCGATAAGTTCGACGCGCAACGATCCGATCTGAAAGATCAGACGCAAGATGCTCC 1648

QY 2156 ggttcagctattgaaatgatggtgtacaaattcaactattttcaactcaaatgatcacg 2215  
 Db 1647 GGTCTGCAACCTTTACATCTGCGGAAAAATCGGTTTATTCTACAC-----TGACTAT 1594

QY 2216 tctgtatacaagttgaatgatcaacgcttgcaccgcaacattaaaccttaattgtat 2275  
 Db 1593 TCGGTTAAACATTACGGCAACAAAGCCTGACACACGCGGAGTAATGTGTCAAAATCT 1534

QY 2276 gatacgggtgttcaatcaagagtggtgataatataatcaagttttgttgagggtgagga 2335  
 Db 1533 GATGACACACTCAAAATCAACGGAGTGAAGATCACAACAAACGATTTTGTACGGAGACGGA 1474

QY 2336 ttctactacaaacttatgaacattcgcaaacggcaaacagatcgtgaaaatgatgatac 2395  
 Db 1473 AAACATATCAGAACGTTACGACAGTTTATCGATGAGGCAATTATACATCCGCGCAAC 1414

QY 2396 tgcttacgtgacccacacggtgttcaattagaataatggtgatcgttattctgtattcga 2455  
 Db 1413 CATACGCTGAGACACCTCTACTAGTTGAAGACAAAGGCCATAATACCTTGTTATTCGAA 1354

QY 2456 gctaatactgggacagaa---gattaccaaagtgaacacaaattataattgggctaac 2512  
 Db 1353 GCCAACACGGGAACAGAAAAACGGATACCAAGGCGAAGATCTTTATTAAACAAAGCGTAC 1294

QY 2513 tatgtggcgatgatgccttcaatataagagttcctcaagcttttgaataataagaag 2572  
 Db 1293 TACGGCGGCGGACGAACTTCTCCGTAAGAAAGGCGAAGGCTTCAGCAGACGCGTAAA 1234

QY 2573 gatcgtgaattggctggttttagtctaattgtgcacttgggtatctttaaagctcactaacaat 2632  
 Db 1233 AAACGGGA---TGCTGAGTTAGCAACGCGGCCCTCGGTATCATAGAGTTAAATATGAT 1177

QY 2633 caaagtaagccaaaggt 2649  
 Db 1176 TACACATTGAAAAAAGT 1160

## RESULT 5

AAA70105

ID AAA70105 standard; DNA; 5940 BP.

XX

AC AAA70105;

XX







[illegible]













GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2002, 05:23:09 ; Search time 12398.7 Seconds  
(without alignments)  
7821.269 Million cell updates/sec

Title: US-09-995-587A-10

Perfect score: 4634

Sequence: 1 gttacaaagacaaatttt.....gttactgttgcaaaagctt 4634

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank:

1: gb\_ba.\*

2: gb\_hgt.\*

3: gb\_in.\*

4: gb\_ov.\*

5: gb\_pat.\*

6: gb\_ph.\*

7: gb\_pl.\*

8: gb\_pr.\*

9: gb\_ro.\*

10: gb\_sts.\*

11: gb\_sy.\*

12: gb\_un.\*

13: gb\_vl.\*

14: gb\_vl.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_on.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vl.\*

30: em\_hgt\_hum.\*

31: em\_hgt\_inv.\*

32: em\_hgt\_other.\*

33: em\_hgt\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
-----						

1	4612.6	99.5	4634	6	AX316622	Sequence
2	637	13.7	2367	6	AX316614	Sequence
3	637	13.7	2394	6	AX316615	Sequence
4	637	13.7	2592	6	AX316616	Sequence
5	562	12.1	4305	1	STRFTF	Sequence
6	459.2	9.9	3600	1	STRFTF	Sequence
7	221.6	4.8	3018	1	LCAJ3194	Sequence
8	136	2.9	18622	1	LCAJ3194	Sequence
9	125	2.9	213190	1	BSUB0004	Sequence
10	102.8	2.2	171798	2	AC007926	Sequence
11	101.8	2.2	155204	2	AC007926	Sequence
12	98.4	2.1	36272	1	STH278471	Sequence
13	98.4	2.1	133075	2	AC099044	Sequence
14	98	2.1	4865	3	AF364131	Sequence
15	96	2.1	189893	2	AC013254	Sequence
16	95.6	2.1	100820	2	AL591365	Sequence
17	95	2.1	584	3	AF201315	Sequence
18	94.8	2.0	10548	3	AE002745	Sequence
19	94.4	2.0	74127	2	AC020125	Sequence
20	94.4	2.0	157141	3	AC016445	Sequence
21	94.4	2.0	177857	3	AC010211	Sequence
22	94	2.0	6987	1	AB047239	Sequence
23	94	2.0	5411	15	AB010124	Sequence
24	92.4	2.0	45636	7	AB044554	Sequence
25	92.4	2.0	539	8	AY022404	Sequence
26	91.8	2.0	953	3	TB5BVSG5	Sequence
27	91.6	2.0	4096	3	AF362374	Sequence
28	90.6	2.0	303650	1	AP001520	Sequence
29	90.2	1.9	303650	1	AP001519	Sequence
30	89.8	1.9	303650	1	AP001515	Sequence
31	89.6	1.9	303249	1	AY023619	Sequence
32	89.4	1.9	494	8	AY023619	Sequence
33	88.6	1.9	1616	3	TBVS117A	Sequence
34	88.4	1.9	166905	3	AC104631	Sequence
35	86.8	1.9	5769	3	AF104350	Sequence
36	86.4	1.9	175978	3	AC010848	Sequence
37	85.8	1.9	2426	8	SDU49822	Sequence
38	84.4	1.8	3149	3	AF019981	Sequence
39	84.4	1.8	3542	8	YSCMTCG16	Sequence
40	84.4	1.8	5931	3	DDU32174	Sequence
41	84.2	1.8	12668	3	AF079445	Sequence
42	84.2	1.8	21451	3	AE003284	Sequence
43	84.2	1.8	104992	2	AC005504	Sequence
44	84.2	1.8	114226	2	AC004710	Sequence
45	84.2	1.8	196490	2	AC005507	Sequence

#### ALIGNMENTS

RESULT 1

LOCUS	AX316622	4634 bp	DNA	linear	PAT 14-DEC-2001
DEFINITION	Sequence 10 from Patent WO0190319.				
ACCESSION	AX316622				
VERSION	AX316622.1	GI:17899731			
KEYWORDS					
SOURCE	Lactobacillus reuteri.				
ORGANISM	Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae; Lactobacillus.				
REFERENCE	1 (sites)				
AUTHORS	van Geel-Schutten,G.H., Rahaoui,H., Dijkhuizen,L. and van Hijum,S.A.				
TITLE	Novel fructosyltransferases				
JOURNAL	Patent: WO 0190319-A 10 29-NOV-2001;				
FEATURES	TNO (NL)				
source	Location/Qualifiers				
	1..4634				
	/organism="Lactobacillus reuteri"				
	/db_xref="taxon:1598"				
RBS	1205..1210				
CDS	1220..3598				

/note="unnamed protein product"

/codon\_start=1

/transl\_table=11

/protein\_id="CAB19335.1"

/db\_xref="GI:17899732"

/translation="MYKVKNAVAATLVASILMGVVTAAHQDVESNNYNGVAEYNT  
ERANGGIGVDGKIIISANSNTTSGSTNOESSANTNTENAVVNESKTNNTNENAVVEN  
KNTNTENAVVENKNTNTENDNSQLKLTNNQPSAAQANLKNNPQAAKAVQNAK  
IDAGSLTDQOIINELKINFSKAEKAKUTFKOLEGNAIVKODPOYAIPIYNAKEI  
KNPATYTDVQGTGRWAHLDSWSPVQDPVTGYVNYMGYQLVIAMGIPNSPTGDN  
HYLLNKYKNDNFDSHRNAGSIFGKETNVFQWEGSALVNDGDTQLFFTSNDSD  
KLRDPHVQLENGDRILVFQANTGDTEDYOSDDOIIYNWANYGGDDAENIKSSFKLLNNK  
KRELAGLAGLGIKLTNNQSKPKVEEYVSLVSTLMACDEVXXKLGDKYLYFSVT  
RVSRGSDRELAKDNTIVGDNVAMIGVSDSLMGKYPKLNNSGVILVITASPNWRAT  
YSYIYAVAGHPQVQLITYSMNKDFASGEGNTATWAPSLVQINPDDTTLVLARATN  
QGWDDSDNNMLGVLEKGAANSALPGEMGKPDVMSLINSRSLGLKPKHPQVQP  
KIDQDPQSGQNTKNTFNGDKPAGKATPDNTNDPSAQPSGQNTNDPSAOSQGO  
NTKNVTPGNEKQGNKDAKQLPQTKNGKSLAGLYAGSLALFGLAAIEKRHA"

BASE COUNT 1530 a 837 c 897 g 1345 t 25 others

ORIGIN

Query Match 99.5%; Score 4612.6; DB 6; Length 4634;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 4634; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	gttaacaaagacaaaattttatataattcttccaaatttaatttcccaactgtaagaacata	60
DB	1	GTTAACAAAGACAAAATTTTATATAATCTTCAAAATTAATTTCCCACTGTAAAGACATA	60
QY	61	aatgggtacctgtttgatgggaataataattgttaactaaaccggcggaacctctttct	120
DB	61	AATGGGTACCTGTTTGTATGGGAATATATATTTGTAACTAACCGCGCGCACCTCTTTCT	120
QY	121	aatgtgctgagatgataatggatgataattactagatggcggtttttatacatataacc	180
DB	121	AATGTGCTGAGATGCTAATATGATGTAATTAATTAATTAATTAATTAATTAATTAATTA	180
QY	181	tcgcagagagaaagaaagcaattagtaggattccatgacgagatgacgagatgacgagat	240
DB	181	TCGCAGGAGAGAAAGCAATTAAGGATCCATATGATGATGATGATGATGATGATGATGATG	240
QY	241	tattagggtcctaaatttcattcaggaactgtaggtactgtatttgggttaaaaggac	300
DB	241	TATTTAGGTCTAAAAATTTCAATTCATGGAACGCTGCTAGGTACTGTATTTGGTTAAAGGAC	300
QY	301	gcaaacccgaagatccatcaatccgcatcttatactggatagaatacaataaagttagcg	360
DB	301	GCAAACCCGAAAGATTCATCAATCCGCAATTTATCTGGATAGAATCAATAAAGTAGCG	360
QY	361	gaattcagatcgcagcaaaaaagtagatcacacaaaaataatgcaaaacggcccatgcta	420
DB	361	GAATTGAGATCGCAGCAAAAAGTAGATACACAAAATATGCAAAACGGCCCATGCTTA	420
QY	421	acaacagacttgcctattatagccacgagtacatacgcgatatgctgatagtcactgca	480
DB	421	ACAACAGCACTTGCCTATTATAGCCACGATGTACATACGGATATGCTGATAGTACCTGCA	480
QY	481	acatcatcgaacgtgagaataatcttccaaatggatatgtttcagaagaagaatatgtg	540
DB	481	ACATCATCGCAACGTGAGAAATATCTTCCAAATGGGATATGTTTCAGAAAAGAAATATTGG	540
QY	541	tcttattcactaatgctaatccccaatgctagtaatttttaatttcgggtattctagaact	600
DB	541	TCATTACTAATGCTAATCCCAATGCTAGTAATTTTAATTTCCGTTATTTAGTAATCT	600
QY	601	aaagttcactatatttggttaaaaaacttttcggtcggtgtgaagtcgattatcggtat	660
DB	601	AAAGTTCACTATATTGGTTTAAAAAATTTTTCGGTTCGGTTCGGTTCGGTTCGGTTCGG	660
QY	661	tcaaacactattattataatttcccttggcgaactgttggtgacaaagccaggaamca	720
DB	661	TAATAAGATTAACTTCTTAAGTCTGCTGAAAAGGGTGCAAAATTTGACCTTTAAGGACTT	720



```

|||||
Db 3961 GSCCATGGTATTTCATTTGGTCTCTTCTCTCTGGTGAATAGGGATCTGGCCAAATAGA 4020
QY 4021 tagtactcttaacgctctcgaatatactcaagccaagaataatcacgccatgatctg 4080
Db 4021 TAGTACTCTCTAAACGCTCTCGAATATCATTCGAAGCCAAGAAATTCACGCCCATGATCTG 4080
QY 4081 gqgcaatgaatgacaaaattcttagaataagaccctaaagagatcaattaagccctgat 4140
Db 4081 GAGTCAATGAATGACAAATTTCTTAGGAATAGACCTTAAGAGATCAATTAAGCCCTGAT 4140
QY 4141 attgaattcggaggaaggagttgtcccaaatgtgcgttataataaccagggttaataac 4200
Db 4141 ATTTGAATTCCGAGAGGGGAGTTGTCCAACATTTGCGCTTAATAATACAGGGTTAATAC 4200
QY 4201 ggcctgggctctacaggttaattgtatttttggctcagatcagtgatagaaaccaca 4260
Db 4201 GGCCCTGGGCTCTACGGTAATATGTATTTTGGCTCAGATCAGTGATAGAAACCCACA 4260
QY 4261 gatttagcttgcgggtgagtgctgctgaagctctcaattacttctgtaaccatgttga 4320
Db 4261 GATTAGCTTGGCGGTGGAGTGCTGTGAAGTCTTCAATTAATCTCGTTACCATGTTGA 4320
QY 4321 ttgctaattcgtatgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4380
Db 4321 TTGCTAATCTGATGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4380
QY 4381 taaggtataacaggaattcttgaactatataatgatactcctccaataataatattagccg 4440
Db 4381 TAAGGTATAACAGGAATTTCTTGTACTATATGATCTTCCAATATATAATATTAGGCG 4440
QY 4441 ataagaaatgaccagctaccatttctgtgcttagtgatgaataataatcgaatagactca 4500
Db 4441 ATAAGAAATGACCAGCTACCAATTTCTGTATGCTTAGTGAATATATTCGGATGATGCTCA 4500
QY 4501 ccctcaacaatccaatttcacggaggtgagtaataatcatcgagagcaggaatgattgg 4560
Db 4501 CCCCTCAACAAATCCAAATTTACGGAGGTGAGTAATCATCGCAGAGCTAGGAATGATTGG 4560
QY 4561 aggaacgaacaggtccatcgcgagtgctgctattgttgatttagcacaagcagcgttact 4620
Db 4561 AGGAACGAACAGGTCCATCGGCGCAGTGTGTTGGATTTTAGCCAAAGCAGCGTTACT 4620
QY 4621 gcttgcataaagctt 4634
Db 4621 GCTTGCAAAAGCTT 4634

RESULT 2
AX316614
LOCUS AX316614
DEFINITION Sequence 2 from Patent WO0190319.
ACCESSION AX316614
VERSION AX316614.1 GI:17899728
KEYWORDS
SOURCE Lactobacillus reuteri.
ORGANISM Lactobacillus reuteri.
REFERENCE Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
AUTHORS van Geel-Schutten,G.H., Rahaoui,H., Dijkhuizen,L. and van
Hijum,S.A.
TITLE Novel fructosyltransferases
JOURNAL Patent: WO 0190319-A 2 29-NOV-2001;
TNO (NL) Location/Qualifiers
1. -2367
/organism="Lactobacillus reuteri"
/db_xref="taxon:1598"
BASE COUNT 814 a 433 c 459 g 661 t
ORIGIN

```

```

Query Match 13.7%; Score 637; DB 6; Length 2367;
Best Local Similarity 59.0%; Pred No. 3.5e-99;
Matches 1266; Conservative 0; Mismatches 821; Indels 60; Gaps 8;

QY 1273 ttttaattggaggggtgttaaccgctcatgctgatacaagtagaagaatacaataatcaacagg 1332
Db 114 TTCTACTGTACAAAGTTACACAGAGGTGATAATGATATTCGTGTTAAAGCTGTGACACTTGG 173
QY 1333 tgttgctgaagttaataactgaacgtcaagctaaatggtcacaattggcgtagatggaataat 1392
Db 174 TAGTGGTCAAGTTAGTGCAGCTAGTGATACGACTATTAGAACCTTCTGCTAATGCAAAATAG 233
QY 1393 tattagtctaacagtagtaatacaaccagtggtcgacaaatcaagaatcatctgctactaa 1452
Db 234 TGTCTTCTTCTGCCCTAATACACAAAATTTCAACAGTCAAGTAGCAAGTTCTGCTGCAAT 293
QY 1453 ca---atactgaaatgctgtgtttaatgaagcaaaaataactaacaatactgaaaaatgc 1509
Db 294 AACATCATCTACAAGTTCCCGCAGCTTCATTAAATAACACAGATAGTAAAGCGGCTCAAGA 353
QY 1510 tgttgtaataagaacaaaataactaacaatactgaaaaatgctgttgaatgaaacaa 1569
Db 354 AAATACTAATACACCCAAAATGATGACGCCAAAAGCTGCACCAGCTAAGCAATCTTC 413
QY 1570 aaatactaacaacacagaaacgataa---tagtcaattaaagttaactaataatgaaca 1626
Db 414 TGAAGCTTAAATAATGAACACAGCTGTAAAGCTTAATGATTTTCACGCTGCAAAAATGATGA 473
QY 1627 accatcagcgcctactcaagcaaaactgaagaagctcaaatcctcaagctcgaagctgt 1686
Db 474 TCACAAATCCAGTAAAGAAATACTACCGCTAAAGTTAAACAAGGATGCTGAAAAACGTTGT 533
QY 1687 tcaaaatgccaaagattgattgctgtagtttaacagatgatacaaatgaataataataa 1746
Db 534 AAAAAAGCGGGAATGATCTCTACACAGTTTAACGTGATGACCAAGATTAAAGCATTAATAA 593
QY 1747 gattaactctctaaagtctgctgaaaaaggtgcaaaattgaccttaaggaacttagaggg 1806
Db 594 GATGAACCTTCGAAAGCTGCAAGTCTGTGTACACAATGACTTATATGATTATTTCCAAA 653
QY 1807 gattggttaagtattgttaagcaagatcccaaatatgctattccttattcctaagtcaa 1866
Db 654 GATTGCTGATACGTTAATCAACAAGATGGTGGTACACAGTTCCATCTTTAAAGCAAG 713
QY 1867 ggaatacaagaatatgcctgcaacatacatcgttagatgccccaaacaggtgaagatggtca 1926
Db 714 TGAATCAAAAATATGCTGCCCTACAACTAAAGATGCACAAACTAATACTATTGAACC 773
QY 1927 tcttgatgctgggactcttggccagtagaagatcctgctcacagggttatgtatataatta 1986
Db 774 TTTAGATGATGGGATTCATGGCCAGTTCAAGATGTTTCGACAGGACAAAGTCTAATTG 833
QY 1987 catgggttatcaactagttattgctatgattgggtattcccaaatcgcgaactggagataa 2046
Db 834 GAATGGCTATCAACTTGTCTATCGCAATGATGGGAATTTCCAAA-----CCAAAATGATAA 887
QY 2047 tcatactatctctttaaacaagaatggtgataatgacttttctcatcttgctgagcgaatgc 2106
Db 888 TCATATCTATCTTATATAAATGAATGATGGTGAATAATGAATTAAAGTCAATTTGGAAGAATGT 947
QY 2107 aggttcaactctttggaaactaaagaacaaatggttcccaagaatgctcaggttcagctat 2166
Db 948 AGGTCCAATTTTGGCTATTAATTTCTACCGCGGTTTCAAGAAGATGGTCAGGATCAGCTGT 1007
QY 2167 tgtaaatgatgagtggtacaaattcaactattttcacctcaaatgatactgctgattacaa 2226
Db 1008 TTTGAACAGTGAATACTCTATCCAAATTTATTTATACAAGGGTAGACACGCTCTGATACAA 1067
QY 2227 gttgaatgatcaacgcttgcacgaacattaaaccttaattgtgtgataaacggtgt 2286
Db 1068 TACCATCATCAAAAATTTGCTAGCGCTACTCTTTTATTAACATGATAATAATGGAATGT 1127
QY 2287 ttcaatcaagagtggtgataattatcaagtttgttgaagggtgatttcaactacca 2346

```

```
Db 1128 ATCACTCGCTCAGGTACGAAATGACTATATATTTTAAAGGTGATGGCTATTACTACCA 1187
Qy 2347 aacttatgaacaattcgcaacggaagatgtaaaatgatgattactgtctacgtga 2406
Db 1188 AACTTTATGATCAATGAAAGCTACTACAAAGGTG--CCGATATATTTGCAATCGGTGA 1244
Qy 2407 cccacagcttgtaaatagaaaatggatcggttatcttctgtaacgtaactggaactgg 2466
Db 1245 TGCTCATGTAATTTGAAGATGTTAATTTGGTATCGGTACCTGTTGTTTGAAGCAAGTACTGG 1304
Qy 2467 gacagaagattaccagaagtcagcaccacaattataattgggctaactatggtggcgatga 2526
Db 1305 TTTTGAAAATATCAAGCGGAGGACCAATTTATTAACCTGGTTAAATTTATGCGGAGATGA 1364
Qy 2527 tgcctcaataaagagttccttcaagcttttgaataataagaaggatcgtaattggc 2586
Db 1365 CGCATTTAATATCAGAGCTTATTAGAAATTTCTTTCCAATGATGATATTAAGAGTCGGGC 1424
Qy 2587 tgggttagctaatggtgcacttggtatcttaagctcaactacaatacaagaagcaca 2646
Db 1425 AACTTGGGCTAATGCAGCTATCGGTATCCTCAACTAATAAGGACGAAAGAAATCTCTAA 1484
Qy 2647 ggttgaagaagtatactcaccattggttatctacttggcttgctgcatgaggta----- 2701
Db 1485 GGTGCACAGTTATATCAACCATTAATTTCTGCACCAATGGTAAAGCATGAATTTGACCG 1544
Qy 2702 -----nnnnnnnaagcttggatgaagtatatctcttctccgtaactcgtaagtcg 2754
Db 1545 ACCAAATAGTAAATTTAGGTAAATAATATTACTTATTTCCGCTACCCGTTTAAATCG 1604
Qy 2755 tgggttcgatcgtaattaaacgctagagataacacaactcggttgataacggttgcat 2814
Db 1605 AGGAAGTAAATGATGCTTGGATGAATGCTAATTAATGCGCTGGTGATATGTTGCAAT 1664
Qy 2815 gattggttacgttccgatagcttaattgggtaagtaacagcattaaataactcaggtgt 2874
Db 1665 GGTGGATATGCTGATAGTCTAACTGGATCTTATAAGCCATTAATGATTCTGGAGT 1724
Qy 2875 cgtattaaactgacagtaacctgcaactggcgactgactacttacttactactgcaagt 2934
Db 1725 AGCTTGACTGCTTGTGCTGCAAACTGGCGACAGCAACTTATTTCATATTATGCTGT 1784
Qy 2935 acctgtagctggtcatctctgcatcaagtattacttacttctacatgagtaacaggaactt 2994
Db 1785 CCGCGTTGCCGGAAGAGTAGCAGGATTAATGATTCTCATATATGACTAATAGAAATGG 1844
Qy 2995 tgcctcgggtgaaggaactatgcaacttggcgaccagttcttagtacaaatcaatcc 3054
Db 1845 AGTAGCGGGTAAAGGAATGGATTCACTTGGGCACCGAGTTCTTACTACAAATTAACCC 1904
Qy 3055 agatgacacgacaactgattagcagctgcaactaacaagtgactgggtggtgggacga 3114
Db 1905 GGATAACACAACTACTGTTTGTGCTAAATGACTAATCAAGGGATGGATTGGGATGA 1964
Qy 3115 ctctagtcggaacgataaatgctcgtggttctttaaagaagtgtagtaacagtcgcgc 3174
Db 1965 TTCAAGCGGAAATCTTGATATGATTGTTGATTT-----AGACTCCGCTGC 2009
Qy 3175 cttaacagggtgaatggggtgaagccagtgactgaggttattgattacaagaagtcctggtc 3234
Db 2010 TTTTACCTGGCGAAGCTGATAAACCCTGTTGATTGGGACTTAAT-----TGGTTA 2057
Qy 3235 aggtctaaagcctcatcaaccagttcaaccacaagattgata-----acctgataca 3288
Db 2058 TGGATTAAACCCGCTGATCTGCTACACCAATGATCTGAAACGCCCACTACACCAGA 2117
Qy 3289 acctctggttcaaacactaagaatgtaacaccagggttaattgggtgataagcctgctgtaa 3348
Db 2118 AACCCCTGAGACACCTAATCTCCCAAAACACCAAGACCTCTGAAATCTCTGGGACACC 2177
Qy 3349 ggaactccctgataaacactaataattgataccaagtgcacaacctctg 3395
```

Db 2178 TCAAACTCCTAATACACCTAATACTCCGAAATTCCTTTAACTCCAG 2224

## RESULT 3

AX316615 AX316615 2394 bp DNA linear PAT 14-DEC-2001  
LOCUS Sequence 3 from Patent WO0190319.  
DEFINITION ACCESSION AX316615  
VERSION AX316615.1 GI:17899729  
KEYWORDS  
SOURCE Lactobacillus reuteri.  
ORGANISM Lactobacillus reuteri  
Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
Lactobacillus.  
REFERENCE 1 (sites)  
AUTHORS van Geel-Schutten, G.H., Rahaoui, H., Dijkhuizen, L. and van  
Hijum, S.A.  
TITLE Novel fructosyltransferases  
JOURNAL Patent: WO 0190319-A 3 29-NOV-2001;  
TNO (NL)  
FEATURES  
source Location/Qualifiers  
1..2394  
/organism="Lactobacillus reuteri"  
/db\_xref="taxon:1598"  
BASE COUNT 829 a 437 c 464 g 664 t  
ORIGIN

Query Match 13.7%; Score 637; DB 6; Length 2394;  
Best Local Similarity 59.0%; Pred. No. 3.4e-99;  
Matches 1266; Conservative 0; Mismatches 821; Indels 60; Gaps 8;  
Qy 1273 tttaattggaggggttgaaccgctcactgcatcaagtagaagaaatacaaatcaaacgg 1332  
Db 141 TTCCTACTGTACAAAGTTACAAACAGGTGATAATGATATGCTGTGTTAAAGTGTGCACATTGG 200  
Qy 1333 tggctggaagttaactgacgtaacgtaagctaatggctgaatggcgtagatggaaaaat 1392  
Db 201 TAGTGGTCAAGTTAGTGCAGCTAGTACGACTATTAGAACTTCTGCTTAATGCAATAG 260  
Qy 1393 tattagtgctaacagtaatacaaccagtggtcgcaaatcaaatcaaatcactgactactaa 1452  
Db 261 TGTCTTTTCCCGCTAATACACAAATTTCAACAGTCAAGTAGCAAGTCTCTGCTCAAT 320  
Qy 1453 ca---atactgaaaaatgctgtgtttaaataagaagcaaaaaataactaaactgctgaaatgc 1509  
Db 321 AACATCATCTACAAGTTCGCGAGCTTCATTAAATAACACAGATAGTAAGCGCTCAAGA 380  
Qy 1510 tgttgtaatagaacaaataactaaacaatactgaaatgctgtgttgaatgaacaa 1569  
Db 381 AAATACTAATACGCAAAAATGATGACAGCAAAAAGCTGCACCAGTCAACGAAATCTTC 440  
Qy 1570 aaatactaacacacagaaaacgataa---tagtcaattaaagttaactaaatgaacaa 1626  
Db 441 TGAAGCTAAAAATGAACCCAGCTGTAAACGTTAATGATCTTTCAGCTGCACAAAATGATGA 500  
Qy 1627 acctacagcgtactcaagcaaaccttgaagaagtaaatcctcaagctgtaaggtctg 1686  
Db 501 TCAACAATCCAGTAAAAAGAAATACTACCGCTAAGTTAAACAAGGATGCTGAAACGTTGT 560  
Qy 1687 tcaaatgccaagattgactgcggtagtttaacagatgatacaaatgaatgaatgaataa 1746  
Db 561 AAAAAAGCGGGAATGTATCTTACAGTTTAACTGATGACCAAGATTAAGCAATTAATAA 620  
Qy 1747 gattaaactctctagctgctggaagaagggtgcaaaattgaccttaagcacttagaggg 1806  
Db 621 GATGAACCTTCGAAAGCTGCAAAAGCTGGTACACAAATGACTTATATGATTTCCTCAAAA 680  
Qy 1807 gattggtaattgctattgttaagcaagataccacaatgctgattctcttattcttaata 1866  
Db 681 GATTGCTGATACGTTAATCAACAAGATGGTCCGGTACACAGTTCCATTTCTTTAAAGCAAG 740  
Qy 1867 ggaatacaagaatatgctcgtgcaacatacactgtagatgccccaaacaggtgaagctca 1926

```

Db      741  TGAATCAAAATATGCGCTGCGCTACAACTAAAGATGCACAACTAATACTATTGAACC 800
QY      1927  tcttgatgctgggactcttggccagttacaagatcctgtcacaggttatgtatcattaa 1986
Db      801  TTTAGATGTATGGGATTCACTGGCCAGTTCAAGATGTTGCGACAGGACAAGTTGCTAATTG 860
QY      1987  catgggttatacaactagttatgctatgatgggtattccaaattccgcaactggagataa 2046
Db      861  GAATGGCTATCAACTTGTTCATCGCAATGATGGAAATTCCTCAAA-----CCAAATGATA 914
QY      2047  tcaatactatctctttacaacaagtatggtgataatgactttctcattggcgcaatgc 2106
Db      915  TCATATCTATCTTATATATAAATAGTATGCTGATATGAATTAAGTCATTTGGAAGATGT 974
QY      2107  aggttcaactctttgaaactaaagaacaactgtgttccaagaatggtcagggttcagctat 2166
Db      975  AGGTCCAAATTTTGGCTATAATTTCTACCGCGGTTTCAACAAGAAATGCTCAGGATCAGCTGT 1034
QY      2167  tgtaaatgatggtgataaattcaactattttcaacctcaaatgatacgtctgattacaa 2226
Db      1035  TTTGACACGTGATAACTCTATCCAAATTTATTTATACAGGGTAGACACGCTGTGATAACAA 1094
QY      2227  gttgaatgatcaacgccttgctaccgcaacattaaaccttaattgttgatgataacggtgt 2286
Db      1095  TACCAATCATCAAAAATTTGCTAGCGCTACTCTTTATTTAACTGTATTAATGAAGATGT 1154
QY      2287  ttcaataagagtggtgataattatacaagtttgttgaagtgatggatttccactacca 2346
Db      1155  ATCACTCGCTCAGGTACGAAATGACTATATTGTATTGAAAGGTGATGGCTATTACTACCA 1214
QY      2347  aacttatgaacatttcgcaacgcgcaaatcgtgaaatgatgattactgcttaagtga 2406
Db      1215  AACTTATGATCAATGGAAGCTACTACAAAGGTG---CCGATAATATTGCAATGCGTGA 1271
QY      2407  ccacacgtgttcaattagaagaatggtgatcgttatctgttattcgaagcttaactgg 2466
Db      1272  TGCTCATGTAATTGAAGATGGTAATGGTGTATCGGTACCTTGTTTTGAAGCAAGTACTGG 1331
QY      2467  gacagaagattcccaagtgacgacccaattataatgtggcttaactatggtggcgatga 2526
Db      1332  TTTGGAAAATTTCAAGGCGGAGGACCAAAATTTATACTGGTTAAATATTGCGCGAGATGA 1391
QY      2527  tgccctcaatatataagagttcctccaagcttttgataataagaagatcgtgaattggc 2586
Db      1392  CGCATTTATATCAAGAGCTTTATTAGAAATCTTTCCAAATGATGATTAAGAGTGGGG 1451
QY      2587  tggtttagctaagtgtgacttggatctttaaagctcactaacaactcaagtaagccaaa 2646
Db      1452  AACTTGGGCTAATGCAGCTATCGGTATCCTCAAACTAATAAGGACGAAAGAATCCTAA 1511
QY      2647  ggttgaagaagtatactcaaccttggatcctacttggatggcttgcgatgagta----- 2701
Db      1512  GGTGGCAGAGTTATATACTCAACCTAATTTCTGCACCAATGGTAAGCGATGAATTCGAGC 1571
QY      2702  -----nnnnnnaagcttggatgaagattattctcttcctcgtaactcgtgtaagtgc 2754
Db      1572  ACCAAATGTAGTTAAATTAGGTAAATAAATATTACTTTTGGCGGTACCGGTTTAAATCG 1631
QY      2755  tggttccgatcgtgaattaaaccgctaagataacacaaactcgttgggtgataaacgttgc 2814
Db      1632  AGGAAGTATGATGCTTGGATGAATCTAATTAATATGCGGTGGTGGTGAATGTTGCAAT 1691
QY      2815  gattgggttacgttcgcgatagtttaattgggtgaagtacaagccattaaataactcaggtg 2874
Db      1692  GGTGGGATATGTTGCTGATAGTCTAACTCGATCTTATAAGGCCATTAATGATTCCTGAGT 1751
QY      2875  cgtattaaactgcatgaatccttgcgaactggcggtactgctacttacttactactcaggt 2934
Db      1752  AGTCTTGAATGCTGTTCTGTTCCTGCAAACTGGCGGACAGCAACTTATTTCATATTATGCTGT 1811
QY      2935  acctgtagctggtcatccttgatcaagtaattaatcttcttcatgtagtaacaagactt 2994

```

```

Db      1812  CCCCCTTGCCGGAAAAAGATGACCAAGTATTAGTTACTTCATATATGACTAATAGAAATGG 1871
QY      2995  tgcttcagggtgaagaaactatgcacttgccacacttggttcttcttagtacaaatcaatcc 3054
Db      1872  AGTAGCGGGTAAAGGAATGGATTCAACTTTGGCCACCGAGTTCTTACTACAAATTTAACCC 1931
QY      3055  agatgacacgacaaactgtattagcacgtgcaactaacaagaagtgactgggtgctgggacga 3114
Db      1932  GGATAACACAACTACTGTTTGTAGCTAAATGACATTAATCAAGGGATTTGGATTGGGATGA 1991
QY      3115  ctctagtcggaacgataataatgctcgtggttctttaaagaaggtgcagtaacagtcgcgc 3174
Db      1992  TTTCAAGCGAAATCTTGATATGATTTGGTGATTT-----AGACTCGCTGC 2036
QY      3175  ctaccagggtgaatggggtgaagccagttgactggagtttgatatacagaagtcctggctt 3234
Db      2037  TTTACCTGGCGAACCTGATAAACCTGTTGATTTGGACTTAAT-----TGGTTA 2084
QY      3235  aggtttaaagcctcatcaaccagttcaacaaaagattgatca-----acctgatcaaca 3288
Db      2085  TGGATTAAACCCGATGATCTGTACACCAAAATGATCTGAAACGCCCAACTACACCAGA 2144
QY      3289  acctctggtcacaacactaagaatgtcacaccaggtaagtgtgataagcctgctggttaa 3348
Db      2145  AACCCCTGAGACACCTTATACTCCCAAAACACCAAGACTCCTGAAAATCTCTGGGACACC 2204
QY      3349  ggcaactccttgatacaactaattgatccaaagtgcacaacctcttg 3395
Db      2205  TCAAACCTCCTTAATACCTTAATCTATCTCCGGAAATTCCTTTAACTCCAG 2251

RESULT 4
LOCUS      AX316616          2592 bp          DNA          linear          PAT 14-DEC-2001
DEFINITION Sequence 4 from Patent WO0190319.
ACCESSION  AX316616
VERSION     AX316616.1  GI:17899730
KEYWORDS   Lactobacillus reuteri.
SOURCE      Lactobacillus reuteri.
ORGANISM   Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
            Lactobacillus.
REFERENCE  1 (sites)
AUTHORS   van Geel-Schutten,G.H., Rahaoui,H., Dijkhuizen,L. and van
            Hijum,S.A.
TITLE      Novel fructosyltransferases
JOURNAL    Patent: WO 0190319-A 4 29-NOV-2001;
TNO (NL)
FEATURES   Location/Qualifiers
            source          1..2592
                        /organism="Lactobacillus reuteri"
                        /db_xref="taxon:1598"
            misc_signal     1..67
            RBS             29..32
            RBS             54..57
            misc_signal     2438..2592
BASE COUNT 896 a 465 c 501 g 730 t
ORIGIN

```

```

Query Match      13.7%; Score 637; DB 6; Length 2592;
Best Local Similarity 59.0%; Pred. No. 3.4e-99;
Matches 1266; Conservative 0; Mismatches 821; Indels 60; Gaps 8;

QY      1273  ttttaatggggggggttgaaccgctcatcgtgatacaagtagaagaagtaacaattacaacgg 1332
Db      181  TTCTACTGTCAAGTTTACACACAGGTGATGATGATGATGATGATGATGATGATGATGATGATG 240
QY      1333  tgttctggaagttaactgaactgaactgaactgaactgaactgaactgaactgaactgaactga 1392
Db      241  TAGTGTCAAGTTAGTCGACGTAGTACGACTATTAGAACTTCTGCTAATGCAATAG 300

```



QY 1393 tattagtgtaacagtaataacaaccagtggtctcgacaaaatacaagaatacatctgactactaa 1452  
 Db 301 TGTCTTCTTTCGCCGCTTAATAACAAATCTTAACAGTCAAGTAGCAAGTTCTGCTGCAAT 360  
 QY 1453 ca---atactgaaatgctgtgttaataagaagcaaaaataactaacaataactgaaaaatgc 1509  
 Db 361 AACATCATCTACAAAGTTCCGACGCTTCAATTAATAACACAGATAGTAAGGGGCTCAAGA 420  
 QY 1510 tgttgtaatagaacaaaataactaacaatactgaaaaatgctgtgttgaataaacaacaa 1569  
 Db 421 AAATACTAAATACAGCCAAAATGATGACGCAAAAAGCTGCACACAGTAAGCAATCTTC 480  
 QY 1570 aaataactaacaacagagaacagataaa---tagtcaattaaagttaactaataatgaaca 1626  
 Db 481 TGAAGCTAAATAATGAACACGCTGTAAACGTTAATGATTTCTACGCTCAAAAATGATGA 540  
 QY 1627 accatcagccgctactcaagcaaacctgaaagctaaatccctcaagctgtaagcgctgt 1686  
 Db 541 TCAACAAATCCAGTAAATAAGAAATCTACCGCTAAGTTAAACAAGGATGCTGAAAACGTTGT 600  
 QY 1687 tcaaaatgccaaatgattgatgccggttagtttaacagatgatcaaaataatgaataataaa 1746  
 Db 601 AAAAAGCGCGGAATTGATCTTAACACTTTAACTGTAGTACACAGATTAAGCAATTAATAA 660  
 QY 1747 gattaactctctaagctctgaaagaggtgcgaattgacacctttaaggaacttagaggg 1806  
 Db 661 GATCAACTTCTCGAAAGCTGCAAGTCTGGTACACAATGACTTATATGATTTTCCAAAA 720  
 QY 1807 gattgtaatgctattgtaagaagatccacaatatgctattctctattcttaactgtaa 1866  
 Db 721 GATTGCTGATACGTTAATAACAAGATGCTGCGTACACAGTCCCTCTTTTAAAGCAAG 780  
 QY 1867 ggaatacaagaaatgctgcacacatacactgtagatgcacaaacaggtgaagaaggctca 1926  
 Db 781 TGAATCAAAAATATGCTGCCCTACAACTAAAGATGCACAACTAATACTATTGAACC 840  
 QY 1927 tcttgatgctgggaactcttgccagtaacaagatccctgtaacaggttatgtatctaatta 1986  
 Db 841 TTTAGATGTATGGGATTCATGGCCAGTTCAAGATGTTCCGACAGGACAAGTTGCTAATTG 900  
 QY 1987 catgggttatcaactagtattgctatgagtggtgtattccaaattccgcaactggagataa 2046  
 Db 901 GAATGGCTATCAACTTGTCTATCGCAATGTGGGAATTTCCAAA-----CCAAAATGATAA 954  
 QY 2047 tcatactatctcttttacacaagatggtgataatgaactttctcatcttgccgcaaatgc 2106  
 Db 955 TCATATCTATCTCTTATATAATAAGTATGTTGATAATGAATTAAGTCAATTGGAAGAATGT 1014  
 QY 2107 aggttcaactcttggaactaaagaaacaaatggtttccaagaatggctcaggttcagcttat 2166  
 Db 1015 AGGTCCAAATTTTGGCTATATAATTCTACCGCGGTTTCAAGAATGGTCAGGATCAGCTGT 1074  
 QY 2167 tgaatgatgatgggtacaaacttcaactatttcacctcaaatgatatcgctctgattacaa 2226  
 Db 1075 TTTGAACAGTGAATACTTATCCAAATTTATTTATACAGGGTAGACAGCTGTGATAKAA 1134  
 QY 2227 gttgaatgatcaacgctgtctacgcgaacattaaaccttaattgttatgataaacggtgt 2286  
 Db 1135 TACCAATCATCAAAAATTTCTAGCGCTACTCTTTTATTAACTGATAATAATGGAATGT 1194  
 QY 2287 tcaatcaagagtggtgataaattcaagtttgttgtaaggtgatggatttcaactacca 2346  
 Db 1195 ATCACTCGCTCAGGTACGAATGACTATATTGTTTGAAGGTGATGGCTATTACTACCA 1254  
 QY 2347 aactatgaacattcgcaacggcagaagatcgtgaaaaatgatgattactgcttacatga 2406  
 Db 1255 AACTTATGATCAATGGAAAGCTACTACAAAGGTG---CCGATAATATGCAATGCGTGA 1311  
 QY 2407 cccacagttgttcaattagaaaatggtgactgttatcttctgtatcgaaagtaatactgg 2466  
 Db 1312 TGCTCATGTAATTCAAGATGTAATGTGTGATCGGTACCTTGTCTTTTGAAGCAAGTACTGG 1371

QY 2467 gacagaagattaccaaaagtgaacgaccacaaattttataatttgggctaactatggtgggatga 2526  
 Db 1372 TTTGGAAAATTTATCAAGCGGAGGACCAAAATTTTAACCTGGTTAAATTTATGGCGGAGTGA 1431  
 QY 2527 tgcctcaaatattgaagagttctcttcaagcttttgaataataagaaggatcgtgaattggc 2586  
 Db 1432 CGCATTTAATATCAAGAGCTTATTTAGAATCTTTTCCAAATGATGATATTAGAGTCCGGC 1491  
 QY 2587 tggtttagtaattggtgcacttggtaactttaaagctcaatacaatacaagtaagccaaa 2646  
 Db 1492 AACTTGGCTAATGCAGCTATCGGTATCTCTCAAACTAAATAAGACGAAAAAGAACTCTAA 1551  
 QY 2647 ggttgaagaagataactcaccatggtgtatctacttcttgatggtctgcgatgaagta----- 2701  
 Db 1552 GGTGGCAGGTTATACTCACCATTAAATTTCTGCACCAATGGTAAAGCATGAAATTTAGCGG 1611  
 QY 2702 -----nnnnnnaagcttgggtgaagttattatctctctcctcctaactcgtgtaagtcg 2754  
 Db 1612 ACCAAATGTAGTTAAATTTAGGTAAATAATATTACTTTTGGCGCTACCCGTTTAAATCG 1671  
 QY 2755 tggttccgctgctgaattaaaccgctaaagataacacaatcgttgggtgataacogttgctat 2814  
 Db 1672 AGGAAGTAAATGATGATGCTTGGATGAATGCTAATTTATGCGTTGGTGATGAATGTTGCAAT 1731  
 QY 2815 gattgggttaactccgtagacttaattgggtaagtaacagaccataataactcaggggtc 2874  
 Db 1732 GGTGGGATATCTGCTGATAGTCTTAACGGATCTTTATAAGCCATTAATGATTCTGGAGT 1791  
 QY 2875 cgtattaactcatcagtaacctgcgaactggcgtaactactctacttacttactactgcaqt 2934  
 Db 1792 AGTCTTGACTGCTTCTCTCTCGAAACGCGGAGCAACTTATTCATATTATGCTGT 1851  
 QY 2935 acctgtagctggtcatcctgctcaagttattatctcttcatcagtagtaacaagagact 2994  
 Db 1852 CCCCCTGCCGGAAGATGACCAAGTATTAGTTACTTATATATGACTAATAGAAATGG 1911  
 QY 2995 tgettcaagtgaaagaaactatgcaacttgggcaccaagttctcttagtaacaataatcc 3054  
 Db 1912 AGTAGCGGGTAAAGGAATGGATTCAACTTGGCACCGAGTTTCTTACTACAAATTAACCC 1971  
 QY 3055 agatgacacgcacaactgtattagcagcgtgcaactaacaagtgactggtggtgggaagca 3114  
 Db 1972 GGATAACACACTACTGTTTGTAGCTAAATGACTAATCAAGGGGATTTGGATTGGGATGA 2031  
 QY 3115 ctctagtcggaacgataataatgctcggtgttctttaagaagtgtagctaaacagtgccgc 3174  
 Db 2032 TTCAAGCGAAAATCTTGATATGATTGGTGATTT-----AGACTCCGCTGC 2076  
 QY 3175 cttaccagtgaaatggggtgaagccagtgactggagtttgattaaacagaagtcctggtc 3234  
 Db 2077 TTTACCTGGCAACGTGATAAACCTGTGTGATTGGGACTTTAAT-----TGGTTA 2124  
 QY 3235 aggttaagcctcatcaaccagttcaaccaaagattgatca-----acctgatcaaca 3288  
 Db 2125 TGGATTAAACCGCATGATCTGCTACACCAATGATCTGAAACCGCAACTACACCAGA 2184  
 QY 3289 acctctggttcacaaacacactaagaatgtcacaccaggttaattggtgataaagcgtggttaa 3348  
 Db 2185 AACCCCTGAGACCCCTAATACTCCCAAAACACCAAGACTCCTGAAAATCTCTGGGACAC 2244  
 QY 3349 ggaactcctgataaactaataatgatcccaagtgacacacactctctg 3395  
 Db 2245 TCAAACTCTTAATACACCTAATCTCCGGAATTTCTTTAACTCCAG 2291

RESULT 5

STRFTF	STRFTF	4305 bp	DNA	linear	BCT 15-FEB-1996
LOCUS	S.mutans	fructosyltransferase gene, complete cds.			
DEFINITION	M18954				
ACCESSION	M18954.1	GI:153635			
VERSION					
KEYWORDS	fructosyltransferase.				
SOURCE	Streptococcus mutans (strain GS5) (clone: pSS22 and pTS102.) DNA.				



ORGANISM	Streptococcus mutans
REFERENCE	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Streptococcus.
AUTHORS	1 (bases 1 to 4305)
TITLE	Shiroza.T. and Kuramitsu.H.K
JOURNAL	Sequence analysis of the Streptococcus mutans fructosyltransferase
MEDLINE	gene and flanking regions
COMMENT	J. Bacteriol. 170 (2), 810-816 (1988)
FEATURES	88115184
source	Draft entry and printed copy of sequence for [1] kindly provided by H.K.Kuramitsu, 06-MAR-1988.
	Location/Qualifiers
	1..4305
	/organism="Streptococcus mutans"
	/strain="GS5"
	/db_xref="taxon:1309"
	/clone="pSS22 and pTS102."
	26..394
	/note="ORF1; putative"
	/codon_start=1
	/transl_table=11
	/product="unknown protein"
	/protein_id="AA88583.1"
	/db_xref="GI:1196925"
	/translation="MESRILEEKKVYRPEFINRIDEKVVFNHNGQEDIRHVVKIMVA
	PLTAHLADGIGITLKQFQSAKLHALAGYDAEMGARPLRRTLOTEVEDKLAELILSGK
	ALSQOTLKIGIQKEKLFDIV"
sig_peptide	681..782
CDS	681..3074
	/note="precursor"
	/codon_start=1
	/transl_table=11
	/product="fructosyltransferase"
	/protein_id="AA88584.1"
	/db_xref="GI:153636"
	/translation="METKVRKKMKKGFYVATTTAMLTGIGLSSVQADENSTQV
	SSLELRQVQENTASSAAENQAKTEVOETPSTPAAATVENTDQTTKVTIDNAAV
	ESKASTKPOAATVTKTAASTPEVGQTNKEDKAKATKEADITTPKNTIDEYGLTEQAR
	KIATEGILNLSLTQKQVEALNKLKLTSDAQTKQHTQYQEFQKIAQTLIAQDERVAYP
	YFNAKAIKNKKAATTRDAQTQGTADLDVWDSWPQDAKTGEVINWNGYOLVWAMGPIP
	TDNHNHLYLKNKYGNDFDHNKAGSIFGYNETPLTQEMSGSATVNEGSGSLQFLFTK
	VDFSDKSNQRLATATVNLGDDQDVRLSVENDKVLTPPEGVMAHYQSYQQRSTFP
	TGADNTAMDPHYVDENGGDVLVFEASTGTENIQEGDIYNFTNGSSSAYNKKSLF
	RFUDDQMDYRNSAWANAAGILKLGDKKTPVEDQFYTPLLSSTWYSDLERPNVYKL
	GDKYLYFTASRLNHGNSNDANKANEVGDVNMVGLYGVSDQLTNGYPLNLSQVPLTA
	SVADPNTATYSYVAVVAGSSDTLLMTAYMTNRNEVAGKGNSTWAPSLFIQVLPDG
	TTKVLAEQOGQWIDWEPSTRDTVTGTLDTAYLPCENDGYIDWNVIGYGLKPHTPG
	QVQTPSPPIHDDIILSPENSFDDGHLVLPKPKVNNDSAGRIDQSRNMSGSLNVAENP
	SAGNISVAPSSQKSIWNNTKETKAHVSTTEKKQKGNPSFAALLALFSAFCVSIQKPV
mat_peptide	783..3071
	/product="fructosyltransferase"
terminator	3150..3179
	/note="putative"
CDS	complement(3265..3951)
	/note="ORF3; putative"
	/codon_start=1
	/transl_table=11
	/product="unknown protein"
	/protein_id="AA88585.1"
	/db_xref="GI:1196926"
	/translation="MFSSQKLKERRKKLGLSQAQTADKLGISRPSPFNWENIGKTKPNQ
	KNLDKLAHLKVDVSAVFLSQDHIDEIYIRLNSNEITKLYKSHLQEQDQKKRLNKLK
	KRPYRVKRLKSAGTGYSGFGDNCFTDVFYDEIDHDFASWTFGDSMEFIPLFNGEAL
	IKOTGFDYGAIVADWDGOTYIKKYVREETGLRLVSLNKKYADKFAFYDENPRIIGL
	IVGNFIPLSG"
CDS	4269..>4305
	/note="ORF2; putative"
	/codon_start=1
	/transl_table=11
	/product="unknown protein"
	/protein_id="AA88586.1"
	/db_xref="GI:1196927"
	/translation="MANKKIRIRLKA"

Db	1513	CTAAATGATAATCATATTATCTCTTTATATAATAATATGGAGATAATAATTTTGATCATTT	1572
Qy	2097	ggcgcgaatgcagggttcaatcttttggaactaaagaacaaatgcttccaaagaatggttcag	2156
Db	1573	GGAAAATATGCGAGGTTCTATCTTTGGTTATATGAAACACCCCTAACTCAAGHAATGGTCAG	1632
Qy	2157	gttcagctattgtaaatgatggtgacaattcaactattttcacctcaaatgatatacgt	2216
Db	1633	GTTCAGCTACCGTAATGAAGTGGCAAGTTTGCAGTTATTCTTCAACCAAGGTTGATACTA	1692
Qy	2217	ctgattacaagttgaatgatcaacgccttgctaccgcgaacattaaaaccttaattggtgatg	2276
Db	1693	GTGACAAAACAGTAGTAACAATCAACGCTTTTACCAACAGCGAGCTGTAAATCTTTGGCTTTGATG	1752
Qy	2277	ataacgggtgttcaatcaagagtggtgataattatcaagttt---gtttgaaggtgatg	2333
Db	1753	ACCAAGATGTTAGAATTTCTTCTGTTGNAATGATAAAGTTTTTAAGCCCTGAAGGCGTGA	1812
Qy	2334	gatttcactaccaaacttatgaacaaattcgcaacggcgaagatcgtaaaatgatgatt	2393
Db	1813	TGCGCTATCATTTATCAAAAGTTATCAACAAATGGCGCTTCAACCTTTACAGGTGCTGATAATA	1872
Qy	2394	actgcttaactgtagccacacagttgttcaatttagaataatggtgatcgttatcttctgtattcg	2453
Db	1873	TTGCTATGGCTGATCCACATGTCATTTGAAGATGAGAATGGAGATCGCTATCTTGTCTTTG	1932
Qy	2454	aagctaatactgggcagagaagattaccaaagtgcagaccaaaattataatctgggctcaact	2513
Db	1933	AGGCTAGTACAGGTACAGAGAAATATCAAGGTGAAGATCATAGATTTACACTTTACTTACT	1992
Qy	2514	atggtggcgatgatgccttcaataataagagtctcctcaagcttttgaataataagaaggg	2573
Db	1993	ATCGCGCGAGCTCTGCTTATATATGTTAAAGTCTTTTTAGATTTTAGATGATCAAGATA	2052
Qy	2574	atcgtgaaattggtggttttagctaatagtgacattggtgacattttaaagctcaactaaacatc	2633
Db	2053	TGTATAACCGTGCACGCTGGGCCAATGCAGCTATTTGGTATTTTAAAACTTTAAGGGCGATA	2112
Qy	2634	aaagtaagccaaagttgaagaagtatactcaccattggtatctacttctgattggtgcttgcg	2693
Db	2113	AAAAACACCTGAGGTAGATCATTTTTACAGCCCTTTACTAAGTTCAACAATGGTTTCGG	2172
Qy	2694	atgaggt-----annnnnaagcttgggtagaagtatctctctctctctctcctcgttaa	2741
Db	2173	ATGNACTGAGCGACCCCAATGTGGTTAAATTAGGAGATAAGTACTATCTTTTTACAGCTT	2232
Qy	2742	ctcgtgtaagtcgtggtccgatactgtaattaacgcgttaagataaacacaatcgttgggtg	2801
Db	2233	CACGCTCTTAATCACGGAAGTAACAACGATGCTTTGGAAATAAAGCAAAATGAAGTTGTTGGTG	2292
Qy	2802	ataacgttgcatagtgttagctttccgatagtcctaagtgggtaagtaagaagccattaa	2861
Db	2293	ATAATGTGGTTATGCTAGGTTATGTTCTTGATCAATTTGACTTAACGGCTACAAACCCCTAA	2352
Qy	2862	ataactcaggtgctgtattaaactgcatcagtcacgtcacctgcgaactggcgctactgtactatt	2921
Db	2353	ATAATAGTGGTGGTTTTAACAGCTTTCAGTTCCAGCAGATTTGGCAACCGCGACCTTACT	2412
Qy	2922	cctactatgcagtacactgtagctggtcatcctctgatacaagtattacttcttcaatga	2981
Db	2413	CTTATTATGCTTCCAGTAGCTGGGATCATCTGATACCCCTGTTAATGACGGCTTATATGA	2472
Qy	2982	gtaacaaggacttgcctcaggtggaaggaaaactatgcaactggcaccaggttctcttag	3041
Db	2473	CCAATCGTAATGAAGTCGCAAGGTAAAGGTAGAATTAACCTGGCGACCTAGTTTCTCGA	2532
Qy	3042	tacaaatacaatccagatgacacgacaaactgtattagcacgtgcacactaaccaaggtgact	3101
Db	2533	TTACGGTTTTACAGATGGGACTACAAAAGTCTTAGCAGAAATGACACAACAGGGTGATT	2592
Qy	3102	gggtgtgggacgactctagtcgga	3125

Db 2593 GGATTGGGATGAACCAAGTCGCA 2616

RESULT 6

STRFTFB 3600 bp DNA linear BCT 12-FEB-2001

LOCUS Streptococcus salivarius fructosyltransferase (ftf) gene, complete

DEFINITION cds; uracil phosphoribosyltransferase gene, 5' end; IS1161

transposase gene, 3' end.

L08445

L08445.1 GI:348000

Streptococcus salivarius.

Streptococcus salivarius

Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

Streptococcus.

1 (bases 1 to 3600)

Rathsam,C., Giffard,P.M. and Jacques,N.A.

The cell-bound fructosyltransferase of *Streptococcus salivarius*:

the carboxyl terminus specifies attachment in a *Streptococcus*

*gordonii* model system

J. Bacteriol. 175 (14), 4520-4527 (1993)

93322332

2 (bases 1 to 3600)

Jacques,N.A.

Direct Submission

Submitted (04-JAN-1993) Institute of Dental Research, Australia

Location/Qualifiers

1..3600

/organism="Streptococcus salivarius"

/db\_xref="taxon:1304"

/tissue\_lib="lambda-A39/ATCC 25975"

<1..56

/standard\_name="IS1161"

/function="transposition"

/note="ORF4; putative"

/codon\_start=3

/transl\_table=11

/product="transposase"

/protein\_id="AAA71924.1"

/db\_xref="GI:348001"

/translation="IHGYSAKKLFLTQTA"

220..3129

/gene="ftf"

220..3129

/gene="ftf"

/function="sucrose cleavage and fructan synthesis"

/codon\_start=1

/transl\_table=11

/evidence=experimental

/product="fructosyltransferase"

/protein\_id="AAA71925.1"

/db\_xref="GI:348002"

/translation="MDIIVNSQSNVTAPKQAECKKMRYSIRKVATVGCATSAIYGLTIAF

LGATQVKAQDVETAPATATATATPTSTASLTVASETAATSVATSEAVSSVAHSEVFA

TKPVTQDPQNTTSSVBEKASTVTWSSDATTATPETAATVAATVAATVAATVAATVAATVA

TASSAADTHTEVDLKVSENAANANLKLNGRIKSIYEENMTSDQIVALTETEIKAL

NKVDFTSDDAIKGTGTLNRYNLKQIVASFLQKSLAVPYFKADTIINMPAFNTVDAQ

TMKKEIDWDSWPQDAKSGVSNWNGYOLVISMAGPNKNSHIVLYLRKYGDNDF

THKNAGPIFGVNALEDQQSGNSYNTSDSGIOLYTKNDTSGGLNWOOLASATIN

LAVENPVVKSVDENHILFGDNTHYQSYPKFMSTFDDDHNDGCDPDTDNYCLDRDP

HIENDGSRILFTSGDNTGDENYQGEKQIYKWSYGGDDAFNLKFLINVKHNLINLA

SWANGSIGLTKLDDNEKPSYAEALPTLTHMTDEVERSVYKMGCKYLLFTASRI

NKSTDAEGTAAAREAVGDVVMGLFVSDSLRGKYPRLNGSLVATSPADMTSTYS

YAVPVEGSDPLLYTSYMTNRGGTAGAENKSTWAPSLIKMNADDTTEGLPKMTNGH

DMWDKSSSLVHVGGQNSAKLPNEDFNDVYVAVSGYGLKPHITYPTVDGSGTGVSEAHG

VLTVTQDKGDKADKPTPVSPTEGNHSVDKTNPKGTSPKADNNOQPSADKEDKPTN

PNPDSPARTPPPYGDHSNDNNSNDHHVAVPKPVTGSDGVRDRRPAQAAEATATPV

PKTIVATGTPVTNTVKESVTEAPKPVSEKQVSHGVDRKANETVTKSDSSKGGNN

TKVAAKLATTPKTPSDSEGSNSNLSILATIFAATIASLALLGYGLVTGKIHLPKK"

3219..>3600

/EC\_number="2.4.2.9"

/function="formation of uridine 5'-monophosphate from

CDS

uracil"  
 /note="ORF2; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="uracil phosphoribosyltransferase"  
 /protein\_id="AA071926.2"  
 /db\_xref="GI:12597937"  
 /translation="MGRQVISHPLIOHKLIRREDTSTKDFRELINELMNGVEV  
 SRDPLEEVEIQTPITKRVQKLSKKLAIVFILRAGIGWDGFLSLVPAKVGHIGM  
 YRDELTPEVYLKLPEDIDQRQI"  
 BASE COUNT 1178 a 719 c 734 g 969 t  
 ORIGIN

Query Match 9.9%; Score 459.2; DB 1; Length 3600;  
 Best Local Similarity 58.1%; Pred. No. 7e-69;  
 Matches 950; Conservative 0; Mismatches 624; Indels 60; Gaps 6;

QY 1539 atactgaaatgctgtgttaataaagaaacaaataactaacaacacagaaacgataata 1598  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 665 ATACATCAGAGGCTGAGTAGAGCACCACCTTCACACAGCAAGCTCAGAAAGCAGCTGATA 724

QY 1599 gtcaattaaagttaactaataatgaacaaccatcagccgctactcaagcaaaacttgaaga 1658  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 725 CACACACTGAAGTAGACCTTAAAGTTTCTGAAACACAGCGGCTAATGCCAACCTTAGTA 784

QY 1659 agctaaatctc-----aagctgctaaagctgttcaaaatgccaaagattgagccggta 1712  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 785 AACTTAATGGTCTATCAAGATCTTATTTGTTGAAGAGAAATATGACCTCAGATCAAAATCGTTG 844

QY 1713 gtttaacagatgataaataatgaataaagataaacttcttaagtctgctgaan 1772  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 845 CCTTGACTGAGGAAGAAATTAAGCACTTAATAAAGTTGACTTTAGTGATGATGCTATCA 904

QY 1773 aggggtgca-----aaatgacctttaaggacttaaggggattggtgaatgctattgtta 1826  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 905 AGGGTACAGGTACTAGCTTGACTTACCGTAACCTTGAAAGATATCGTTGCTAGCTTCTGA 964

QY 1827 agcaagatccacaataatgctattcttcttaataagtaaggaatacaagaataatccgtg 1886  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 965 AGCAAGACAGCAAAATTTGGCAGTTTCTTACTTTAAAGCAGATACAAATCATCAATATGCGTG 1024

QY 1887 caacatacactgtagtgcacaggaagtaagatggtctcatcttgatgctggaactctt 1946  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1025 CCTTCAACTGTTGATGCCCCAACGATGAAGAAAGAAATCGACGTTTGGGATCTT 1084

QY 1947 ggcagataagaatcctgtcacaggttatgtatcttaatacatcaggttatcaactagtta 2006  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1085 GGCCAGTCCAAGATGCTAAGTACAGGTGATTTAGCAACTGGAATGCTTACCAGTTGGTTA 1144

QY 2007 ttgctatgagggatctcaaaatcccaactggaagataatcatctatctctttaca 2066  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1145 TCTCAATGGCCGTGACCAACATAAATAATCA-----AACCAATTTACTTGTCTCTATA 1198

QY 2067 acaagtatggtgataatgactttctcattgcccgaatcaggttcaatctttggaacta 2126  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1199 GAAATATGGGGATATGACTTTTACTTGTGAAATGCGGGTCCCAATCTTTTGGTTTACA 1258

QY 2127 aagaaacaaatggtgttccaagaatggttcaggtttcagctattgtaaatgatggtgataa 2186  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1259 ATGCCCTTGAAGATGACCAACAATGTGAGTTTACGCGACTGTTAACTCTGATGGTAGTA 1318

QY 2187 ttcaactattttcacctcaaatgatacgtctgtattacaagttgaatgatacaacgcttg 2246  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1319 TCCAAGTTTATTATCAAGAAATGATGATGAGTGTGGTGAATTAATACTGGCAACAATGG 1378

QY 2247 ctaccgcaacattaaaccttaattgttgatgataacgggttttcaatcaagagttgtgata 2306  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1379 CTAGTGTACACTTAAGTGGCTGTTGAACGACCAAGTGTGTTATCAAGTCAAGTTGAA 1438

QY 2307 attatcaagttttgttgagggtgattgatttccactaccaaaacttatgaacaattcgcaa 2366  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1439 ATGACCACATCTCTTTGGTGGAGATAATTATCACTATCAAGCTATCAAGCTATCTTAATTTATGA 1498

QY 2367 acggcaagatcgtagaaat-----gatgattactgct 2399  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1499 GTACTTTTGTGATGATGACCATAAACCATGATGTAATCCAGACCGTAGCATTAACACTGTC 1558

QY 2400 tacgtgacccacacgttgttcaattagaaaaatggtgatcgcttatctgtattcgaagta 2459  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1559 TTCTGTGACCCACACATCATTTGA---AGATAACGGTAGCCGTACTCTATTATTTGAATCAA 1615

QY 2460 atactggacagaagattaccaaagtgcagcaccaaatttaattgggctaacatggtg 2519  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1616 ATACAGGTGACGAAACACTACCAGCGGAAACAAATTTTACAATGGTCAAACTATGGAG 1675

QY 2520 gcgatgatgcttcaatattaaagsgtttcttcaagcttttgaataataagaagatcg 2579  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1676 GAGATGACGCGCTTTAAAGCTTAAAGTCAATTCCTCAATATTGTTAAATAAACAACCTTTACA 1735

QY 2580 aattggctgggttttagctaaatggtgcacttgctattctaaagctcactaacaatacaagta 2639  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1736 ACCTTGGATCATGCGCAAAATGTTCTATCGGTATCTTGAACCTTGATGATTAATGAGAAAA 1795

QY 2640 agccaaaggttgaagaagatactcaccacttggttatcttcttctggttcgcatgagg 2699  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1796 ATCCATCTGTAGCTGAACCTTTACACACCGCTTGTCACTAGCCACATGTTACTGATGAAG 1855

QY 2700 ta-----nnnnnaagcttggatgaagttatctcttcttcttcttcttcttcttcttct 2747  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1856 TAGAACCTCCAAAGTGTGTTTAAAGATGGGTGGCAAGTATTACCTCTTCACTGCTTCACGTA 1915

QY 2748 taagtctggttcgactcggtgaattaaaccgctaaggataacacaaatcgttggtgataacg 2807  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1916 TCACAAATCTTACCGATGCTGAAGTACTGTAGCAGCGCGTGAAGCTGTCGTTGATGACG 1975

QY 2808 ttgctatgattggttccgtttccgatgacttaattgggtacaagccatataaataact 2867  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1976 TTGTCATGCTTGGATTGCTATCACACAGCTCCCGAGGAAATATCGTCCACTTAATGGCT 2035

QY 2868 cagggtcgctattaaactgcactgcactgcactgcactgcactgcactgcactgcactgcact 2927  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 2036 CAGGTGTGGTCTTGACAGCTCTGTACAGCTGCTGCGGTGACATCTACCTATTCTTACT 2095

QY 2928 atgcagctactgtagctggtcactcctgaatcaagtatttaattactcttctacatgataca 2987  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 2096 ATGCGGTCTCTGTTGAAGGCTCATCAGATCTCTTTTGGTAACTTCTTACATGACTAACCC 2155

QY 2988 aggactttgctcagggtgaaggaaactatgcaacttggcaccacagtttcttagtacaac 3047  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 2156 GTGGTGAATTTGCCGGTCTGTGAGATAAATCAACATGGGCTCCAGGCTTCCTTATCAAGA 2215

QY 3048 tcaatccagatgcacagacaaacttattagcagctgcacactaaccacaaaggtgactgggtg 3107  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 2216 TGAATGCAGATGATACAGAAAGTTCTTCTTAAGATGACCAATCAAGGTGACTGGATT 2275

QY 3108 ggaacgactctagt 3121  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 2276 GGGATAAATCTAGT 2289

RESULT 7  
 LCAJ3194  
 LOCUS Lactobacillus casei ccpA & tnp genes.  
 DEFINITION AJ003194  
 accession AJ003194  
 version AJ003194.1 GI:2673746  
 keywords catabolite regulator protein; ccpA gene; tnp gene; transposase.  
 source Lactobacillus casei.  
 organism Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 Lactobacillus.  
 reference 1 (bases 1 to 3018)  
 authors Monedero, V.  
 title Direct Submission  
 journal Submitted (04-DEC-1997) Monedero V., Biotechnology, Instituto de



similar to Chlamydia trachomatis Hom1: SwissProt Accession Number P55137

```
/citation=[2]
/codon_start=1
/transl_table=11
/product="yefA"
/protein_id="AAB66480.1"
/db_xref="GI:2318067"
/translation="AKRLRSPADNCGDAEAGSVCL"
/complement(227..1140)
/gene="yefB"
/complement(227..1129)
/gene="yefB"
/citation=[2]
/codon_start=1
/transl_table=11
/product="yefB"
/protein_id="AAB66472.1"
/db_xref="GI:2318059"
/translation="MYNEGVTSPQLARKKNATDKWCGSSIRKILLENPHYGTLOQCR
EYPSVTSKRERSVNLNQIILIEKSHPEIIPLEDPLVQELILTRKRRPQAEVHLFT
NTAVCKCRSMHFKRKGVCSYNGKHLKACSDHVSSEIDLTDKVITNLNLYLK
FSKNYFELSEKALKYKEKIESKIIIEINELNDKRDKNLVSISLANGVISEDYQL
AINTNEDISNFTFLHLSKDLDFQIEKEIIDFKKSLDKFMKEGTLTPEMLHLVD
EIEYHANGTIEINFRFPTVPSA"
1024..1445
/gene="yefC"
1024..1030
/gene="yefC"
1068..1445
/gene="yefC"
```

gene

CDS

gene

RBS

CDS

/note="similar to Enterococcus faecalis plasmid PY117 orf4 encoded by GenBank Accession Number D78257"

/citation=[2]

/codon\_start=1

/transl\_table=11

/product="yefC"

/protein\_id="AAB66473.1"

/db\_xref="GI:2318060"

/translation="MSVAFPLRASCEGLTPTSLYNPLAIAASKPPATYSEKILFTIF
SAPAEPRDMIVERTQSGMKLAKGNPDFREGRPKKFTKQINHALTLLENHSYQVED
MTGLSVSTLVRKKKAAEAING"
complement(1133..1140)
/gene="yefB"
1505..4159
/gene="yefB"
1505..1512
/gene="yefB"
1520..4159
/gene="yefA"
1505..1512
/gene="yefA"
1520..4159
/gene="yefA"

RBS

gene

RBS

CDS

/note="similar to Methanococcus janaschii plasmid purB801 hypothetical protein MJBCS02 encoded by GenBank Accession Number L77119"

/citation=[2]

/codon\_start=1

/transl\_table=11

/product="yefA"

/protein\_id="AAB66474.1"

/db\_xref="GI:2318061"

/translation="MALIDLEDKIAEIVNREDHSDFLYELLVGVDPVPRATIRLKGN
QNLTKRVGEHLKRWKFKKAGKLPDALIDEEQVEYLSAKPRYLIVTDYDGLAK
DTLTKALDYFELPQYDFDLFWAGKIEKFEKENPADIKAAERFARIYDVLRKEN
NIETNRGLDLFLIRLFCFAEDTDFKRNSTNLTKLTTEEDGSNLNKLFDLFTV
LDKNERDDVPSYKPEPVNGQFTDPTHTLEFSAKSKLIIICGELLNKNINPLIF
GSMIOAVASPESSYLGMYHTSVPMNMKVIKPLFLDKLNQSFLDAYDDYTKLENLLTR
IGKIKFDPACGSGNELIITYKELRMEINILRQELLEGEILYPSVTLSQFYGIEI
EDFAHDVAKLSIAEAMNEELKMEVHNARPTLPHTAGDIRCANARIVETVECP
AQGSEEVYFNGPYLGSKQKQKHKSDMSIFGVKNGKMLDYISAFYFGAKYAST
TNKAFVSTNSVTQGOVSIWNLKFKGIQNFANKSPFKWANNKNAIVIVYVG
FGPLDTKNVLYFVDETKLVNSIPVLTGENTLVSSRTKPISDLKPLHFGNPNQD
GGLLFTITEYTDALNKYELAPVYFKFVIGSVFINGLRLCLMNEAKYKIKSNPLI
QERISISKNHREKSTDKGNKALTPWKFPHDTHTNTNYSIVPVSSENFRYPIMPGLA
GADTILSNLIYVIYDAEIIYLLGLMSRMHMTWYKAVAGRLKTDYRSAGLCYNTFPPI

RBS

gene

CDS

```
EISTRKRKEIEAILEILDLRERQGGTAEALYNPSTWPIELKVAHEKLDGIVERAYRQ
KQFESDEERLEVLKLYQEMTER"
4152..4161
/gene="yefB"
4152..5967
/gene="yefB"
4171..5967
/gene="yefB"
/citation=[2]
/codon_start=1
/transl_table=11
/product="yefB"
/protein_id="AAB66475.1"
/db_xref="GI:2318062"
/translation="MENIIEINYNOTGSKKTKNEYGMREMAQAEKRNQYLLVKAP
PASGSRALMFGLDKLINOGLRKVIVAVPERSIGSSFKNTDLKSGFFENKWKYDPRN
NLTVGDSNKSXKSVRFMESDDOVLCTHSTIRFAPEKIDDKAFDNCLLAIDEFHHVS
ADYNSRLGELLRSIIHNSAHVIMTSGSYFRGDSVPILLPDEDELFDKVITYSYEQLD
GIEYLRKGFIGYHYOGQYTSAINEVLDNKKTIHIPNVNNGESTKDKYDEVYKILD
MIGEVEYQDDGTGIIYVKRHSQKTLKIALDVDDQVGRNVNVAIYRNIEALDDLDII
ALGMAKEGDFWPCHEHTLVGYSGLTEIWOIIGRCTRDSYNKTAQFTNLLAMPDAK
DEWVYTVNTMLKAISASLLMQVLTPODFKRRRSEKSTTGELFVKGLKEPSTE
VVKIITIENDLKAKTCKIHKSKTFSGEVDKVLNKVLIPIVKVYPNLSEDELEE
YRQHVLDVTVMKGAKEVVGSGKEFIRMAKDFINIEDLDINLIDSNINPFOKAFEVLSKE
LNSPVKLKIQETIDAKRINFNDEDELVFIWPKV"
6064..6069
/gene="yefC"
6064..7228
/gene="yefC"
6080..7228
/gene="yefC"
/citation=[2]
/codon_start=1
/transl_table=11
/product="yefC"
/protein_id="AAB66476.1"
/db_xref="GI:2318063"
/translation="WASNRHYSINEIMESQFYQITPPKTKQAOYDPEVKFLEII
EFVKNGREPKQVPTDLTERRSLRIGIRKDPDRMEYLKEYDEIGLLEVRKELTIP
KISSDILKLSGLSALLGNDLNNDITRSIFDTSSLOKVTYMPYVAKRKKIKDQKCFE
ELFKKCHKEITGKRRKILTFKNEQDIQNSFYILAGVLLYVEDVGERKAKCTNARL
RCIPENGTESDMLRSLSAELYKHGRVYTDNEDTLLDNVREDDDVSTGTFYVLKSLSTD
PQISSIKNLXKIGFTTGSVENRIRNAENQSTLYAPVEIVTYTQVFNMAKSKFATAIH
HALENNLDVSLGANGKMLVPKFWFVVTLEDLQAVIDEIVMMVHLYD"
complement(7490..8180)
/gene="yefB"
complement(7490..7795)
/gene="yefB"
/citation=[2]
/codon_start=1
/transl_table=11
/product="yefB"
/protein_id="AAB66477.1"
/db_xref="GI:2318064"
/translation="MTVRKELIKQINLITVITKTNQKNPTPWKNILKRYEAEKEFI
LOSSDGKFEEDLSKVRNKLDTLITRAYLESANDYMNPMLEKMHKTEKLKEYDETQTS"
complement(8175..8180)
/gene="yefB"
8277..8298
/note="rho independent terminator; deltaG= -10.2kcal/mol"
/citation=[2]
complement(8308..9470)
/gene="yefB"
complement(8308..9417)
/gene="yefB"
/citation=[2]
/codon_start=1
/transl_table=11
/product="yefB"
/protein_id="AAB66478.1"
/db_xref="GI:2318065"
/translation="MPAGVPPNAVNGVTLKELQKFAKVLKPKYGTSSSSGRRTPAPH
VPPVTYVGEHFAKRSRKKVLKPNLIYTKEGYITTYTNYGRITSVKADLQJGEAKRN
```

terminator

gene

CDS

QY AQTNACKPQDRPDDGGHLLIATQFKGSGQFDNIVPMNSQINRSGKWKWEMEQEW  
 RALSKKPKKVVAVQIEPVYSGDSLRPSYFDVYTKIGSRKEISVSKISLGVRRMETRK  
 MDLYQLIGELKRLNDIIPGEWKIYLYAEVLDDSTWLVFHFRTPENNOIYSDIIPSHV  
 NVSKDIFXTLRELRLFEELRTHRNNDNDVWNLTLTLDRSGEFLDYNVDILAS  
 ELGDIYRTAIWEYNLGLTLPEDDEDPKEFVISVGL"  
 complement(.9464. .9470)  
 /gene="yeef"  
 complement(.9541. .10791)  
 /gene="yeef"  
 complement(.9541. .10791)  
 /gene="yeef"  
 /note="similar to Bacillus subtilis orf94 hypothetical  
 59.7 kDa protein in CWLA-CISA intergenic region: SwissProt  
 Accession Number P45942"

Query Match		2.9%; Score 136; DB 1; Length 18622;
Best Local Similarity		48.6%; Pred. No. 7.3e-14;
Matches	470; Conservative	2; Mismatches 477; Indels 19; Gaps 3;
Qy	55 aacataaatgggtactgtttgatgggaataataataattgttaactaacgcgcgcgcacot	114 
Db	3284 AACATAGTCCCTATTAACTAGCTGGTGAATAATATTCTTGTCTCTAGGACGAGCCA	3343 
Qy	115 ctcttaaatgctcaggtatgataatgataatgataatgataatgataatgataatgata	174 
Db	3344 ATTTGACACTTACCAAACTGCATTTTGGGAACATGCGCTAATGATGGCGGAGGATTACTC	3403 
Qy	175 ttaactcgcaggagagaaaaagcaattagtagaggtccatagcagataaaattatt	234 
Db	3404 TTTACCATTCAGGAATATACAGACCAATCAATAAGTATCTCGTAATTAGTTCCTATTTT	3463 
Qy	235 aggccttattaggtgctaaaaatttcaatgcaactgtagtagtactgattggtta	294 
Db	3464 AAAAGTTTATAGGTTAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC	3523 
Qy	295 aaggacgcaaacccggaagatcatcatcgaatcgccatttatactgtagatagatacaataa	354 
Db	3524 AATGAAGCAAAATATGAAAAATAAAGCTTAATCTCTTAATACAGAAAGAACTCAATT	3583 
Qy	355 gtacgggaattcagatcgcagcaaaaaagtaagaatacacaaaaatacgcaaaagcc	414 
Db	3584 TCAAAAAATCACAGGAAAAATCAACAGATAAAGGAAGCAATAAACTTGCA-----	3634 
Qy	415 atgctaacacacagactgctctattatagccacgagtagacacagatagctgtagta	474 
Db	3635 ---TTAACTCGTGGAACTTTAGAGATFACTCATGAAACAACTAATATTCCTATTTGTT	3691 
Qy	475 cctgcaacatcatcgcagcagtagaataatcttccaatgttgatatgttttcagaaaaaat	534 
Db	3692 CCATCTGTCTCTGAAAAATAGATTTTATATTCCAATGGGATTAGCAGGTGCAGATACA	3751 
Qy	535 attgtgcttattcactaatgctaatcccaatgctagtaatttttaatttcggtattcta	594 
Db	3752 ATTTTGTCAAATCAATTTATGTAATTTACGATGCAGAAATTTACCTTCTAGGCATTTG	3811 
Qy	595 gaatcaaatgtcactatatttggttaaaaaacttttgcgcggtggaagtcgattat	654 
Db	3812 ATGTCCAGGATGCATATACACTTGGGTAAAGCTGTAGCCGGAAGATTAACAACTGACTAT	3871 
Qy	655 cgttattoaacacattatttataataattcccttgcgcgactgtttggtgcaaaagcca	714 
Db	3872 AGGTACTCTGCAGGTTTATGTTATTAACACTTTCCT-----ATTCTGTAATTATCAACA	3925 
Qy	715 ggamaacaccactctcgacactcgctcaaggtatataataactcgaagcttatccca	774 
Db	3926 AGAAGGAGAAATGAAGAACAACTCTTGAAATCTTGATCTTCGTGGAAGACGAC	3985 
Qy	775 gacagctcactggtgctatttatgctcactaacatgcccragttgaaactcgtaagct	834 
Db	3986 GGTGGTACATTAGCAGAGCTTTATAACCTTCACAACTGCCAATTAATTAAGTAGCA	4045 
Qy	835 catgaagcaaatgataaagctgtctttaaagcatatgattgagccctaaagctactgag	894 

Db	4046 CATGAAAGTTGCGTGAATTTAGAAAGCGCTATCGTCAAAACAATTTGTAATCTGAT	4105
Qy	895 caagaataatcgtaagaacatctatttaagatgataaactgactaaagtgaaagataa	954 
Db	4106 GAAGAAGCGCTTAGAAGTATTGCTTAAATTTGTATCAAGAAATGACAGAAGC-GTAGGTGA	4164 
Qy	955 ctttgttaaaccaaatattttataaagacagataaattgttaattgataaaacataattt	1014 
Db	4165 AGGTGATGARAATATATATTGAATTAATTATCAACAACCTGTAAGAGTAAGAAAACA	4224 
Qy	1015 aataaaca 1022	
Db	4225 AATGAATA 4232	

RESULT	9
BSUB0004	BSUB0004
LOCUS	Bacillus subtilis complete genome (section 4 of 21): from 600701 to 813890.
DEFINITION	213190 bp DNA linear BCT 26-NOV-1997
ACCESSION	299107 AL009126
VERSION	299107.1 GI:2632866
KEYWORDS	Bacillus subtilis.
SOURCE	Bacillus subtilis
ORGANISM	Bacillus subtilis
REFERENCE	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus.
AUTHORS	1 (bases 1 to 213190) Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C., Bron, S., Brouillet, S., Brusch, C.V., Caldwell, B., Capuano, V., Carter, N.M., Choi, S.K., Codani, J.J., Connerton, I.F., Cummings, N.J., Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D., Emmerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrali, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Gallier, A., Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Gollightly, E.J., Grandi, G., Guiseppe, G., Guy, B.J., Haga, K., Halech, J., Harwood, C.R., Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hulio, M.F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Klauer-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningsstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H., Masuda, S., Maule, C., Medigue, C., Medina, N., Mellado, R.P., Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M., Portetle, D., Porwollik, S., Prescott, A.M., Presecan, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadale, Y., Sato, T., Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J., Sekowska, A., Seror, S.J., Serrero, P., Shin, B.S., Soldo, B., Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A., Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A., Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitznegger, T., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H.F., Zumstein, E., Yoshikawa, H. and Danchin, A.

TITLE	The complete genome sequence of the gram-positive bacterium Bacillus subtilis
JOURNAL	Nature 390 (6657), 249-256 (1997)
MEDLINE	98044033
REFERENCE	2 (bases 1 to 213190) Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A. Direct submission
AUTHORS	Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur, Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr, adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45 68 89 48
FEATURES	Location/Qualifiers 1. .213190 /organism="Bacillus subtilis"

### Query Match







QY 1297 tcatgctgatcaagtagaagtaacaattacaacgggtgttctgctgaagttaatactgaagc 1356

BASE COUNT	48747 a	31603 c	32329 g	42525 t
ORIGIN				

Query Match	2.2%	Score 101.8;	DB 2;	Length 155204;
Best Local Similarity	50.5%	Pred. No. 3.5e-08;		
Matches 247;	Conservative	0;	Mismatches 242;	Indels 0; Gaps 0;

[illegible]





us-09-995-587a-10.rge

[illegible]

Search completed: September 22, 2002, 05:40:23  
Job time: 21803 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on:      September 26, 2002, 18:26:43 ; Search time 22.4 Seconds
              (without alignments)
              1369.013 Million cell updates/sec

Title:       US-09-995-587A-11
Perfect score: 4149
Sequence:    1 MYKVKCKKNAVATIVGASTIV

```

Scoring table: BLOSUM62

Searched: 105224 seqs. 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Post-processing: Migration Method: 00

Maximum Match 100%  
Listing first 45 summaries

Database : Swissprot 40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query No.	Score	Query %		Length	DB	ID	Description
			Match					
1	1684.5	40.6	797	1	SACB_STRMU	Pl1701 streptococc		
2	1586	38.2	969	1	SACB_STRSL	Q55242 streptococc		
3	751.5	18.1	472	1	SACB_BACAM	P21130 bacillus am		
4	746.5	18.0	473	1	SACB_BACSU	P05655 bacillus su		
5	730.5	17.6	473	1	SACB_BACST	P94468 bacillus st		
6	309	7.4	413	1	INVE_ZYMMO	Q60115 zymomonas m		
7	251.5	6.1	423	1	SACB_ZYMMO	Q60114 zymomonas m		
8	245	5.9	415	1	SACB_RAHQA	O54435 rhaznella aq		
9	242.5	5.8	415	1	SACB_PSESG	O52408 pseudomonas		
10	240.5	5.8	431	1	SACB_PSESH	O68609 pseudomonas		
11	236.5	5.7	415	1	SACB_ERWAM	O43654 erwinia amy		
12	197	4.7	584	1	SACB_ACEDI	Q43998 acetobacter		
13	190.5	4.6	1584	1	KVK1_DICDI	Pl18160 dictyosteli		
14	190	4.6	448	1	AA22_DICDI	Pl14196 dictyosteli		
15	185	4.5	850	1	DEXT_STRMU	Q54443 streptococc		
16	185	4.5	1905	1	TAGB_DICDI	Q34683 dictyosteli		
17	182	4.4	1036	1	HP12_DEIRA	Pl31262 deिनococcus		
18	179.5	4.3	1183	1	CNA_STAAU	Q53654 staphylococ		
19	178	4.3	852	1	CSG_HALHA	P08198 halobacteri		
20	178	4.3	1337	1	DXTF_STRDO	P39653 streptococc		
21	174	4.2	1592	1	GETT_STRDO	P27470 streptococc		
22	172	4.1	1018	1	FNBA_STAAU	Pl14738 staphylococ		
23	171.5	4.1	1365	1	GTF5_STRDO	P29393 streptococc		
24	169	4.1	1296	1	ASAL_ENTFA	P17953 enterococcu		
25	167	4.0	1084	1	HGP3_HAEIN	P44836 haemophilus		
26	166	4.0	1288	1	VACA_HELPJP	Q92kw5 helicobacte		
27	166	4.0	1296	1	VAC1_HELPY	Q48247 helicobacte		
28	164.5	4.0	537	1	ARP_PLAFA	P04931 plasmodium		
29	164	4.0	1291	1	VAC4_HELPY	Q48258 helicobacte		
30	164	4.0	1643	1	OMP6_RICPR	Q53020 r outer mem		
31	163	3.9	2003	1	YDBA_ECOLI	P33666 escherichia		
32	162	3.9	2249	1	OMPA_RICRI	Pl15921 rickettsia		
33	160.5	3.9	1902	1	P2P_LACUC	Pl15293 lactococcus		

```

Query Match      40.6%   Score 1684.5; DB 1; Length 797;
Best Local Similarity 44.3%; pred. No. 1.2e-82;
Matches 370; Conservative 126; Mismatches 230; Indels 109; Gaps 21;

QY    1 MYKGVKNWAVATLVASLTGMSGVVTAHADQOVESNNYNNGVAEVNTERQANGOIGVDGKIIS 60
      ||| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     9 MYKKGKFWVATTITAMLTGLSLVSQADEANS-----TQVSELAERSQV----- 54

QY    61 ANSNTTSGSTNQESSAATNVENAVVNYESKNTNTTENAVVNENKNTN--NTENAVV----- 113
      |||| : ||| | | | : | : | : | : | : | : | : | : | : | : | : | : |

```

```

Query Match      38.2%; Score 1586; DB 1; Length 969;
Best Local Similarity 41.6%; Pred. No. 2.8e-77;
Matches 349; Conservative 133; Mismatches 266; Indels 90; Gaps 18;

Qy      2  YKVGKNWAVATLVASILMGCVTAHADQVDESNNYNGVAEVNTERQANGQIGVDGKIISA 61
      | : | | | : | : | : | : | | : | : | : | : | : | : | : | : | : |
Db      24  YSIRK--VAIVGATSALVCTLAFGLAQYKADQVETAPAVATATATPTSTASLTVAS 80
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Qy      62  NSNITSSTSGESSATNNTENAV--VNESKNNTNNTENAVVNENKNT----- 105
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      81  ETATSVATSEAVSESVAHSEVATKPVETQFQSNTPSPVWEKASSTVVTSSDATTTPSAT 140
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Qy      106 -----NNTENAVVNENKNTNNTENDNSOLKLTNNQPSAATQANLKLNPQAAKAV-Q 157
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      141  VAAVSAPAHTSEAAVEAPTSTASSEAAADTHEVDLKYSENSAANANLSKLNGRKKSIVEE 200
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Qy      158  NAKTD-AGSLTDDOINELNKNESKSAEK--GAKLTFKDLEGIGNALVKDQPVATPYSN 214
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      201  NMTSDQIVALTFEESIKALNKVSDDDAIGTGTSLTYRNLKDIVASELKQDSKLAVPYFK 260
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Qy      215  AKEIKNPPATVYDQAQTKMAHLDDWDSWPQDVPVTGYVSNMGYQLVIAWMGIPNSPTG 274
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      261  ADTIINPPAFNTVYDQAQTKMKKEIDVWDSWPQDAKSGVSNWNGYQLVISMAGAPNK--N 318
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Qy      275  DNHTYILVYNKYGDNDPFSHBNAGSTECTGKFNMYPOEWSGATVNDQCTYQIESTGQWGD 324

```

```

Query Match      38.2%; Score 1586; DB 1; Length 969;
Best Local Similarity 41.6%; Pred. No. 2.8e-77;
Matches 349; Conservative 133; Mismatches 266; Indels 90; Gaps 18;

Qy      2  YKVGKNWAVATLVASILMGCVTAHADQVDESNNYNGVAEVNTERQANGQIGVDGKIISA 61
      | : | | | : | : | : | : | | : | : | : | : | : | : | : | : |
Db      24  YSIRK--VAIVGATSALVCTLAFGLATQVKADQVETAPAVATATATPTETASLTVAS 80
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Qy      62  NSNITSSTGOESSATNNTENAV--VNESKNTNNTENAVVNENKNT----- 105
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      81  ETATSVATSEAVSESVAHSEVATPKVPTETQPSNTTPSVWEKASSTVVTSSDATTPTSAT 140
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Qy      106 -----NNTENAVVNENKNTNNTENDNSOLKLTNNQPSAATQANLKLNPQAAKAV-Q 157
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      141  VAAVSAPAHTSEAAVEAPTSTASSEAAADTHEVDLKYSENSAANANLSKLNGRKISVEE 200
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Qy      158  NAKTD-AGSLTDDOINELNKNESKSAEK--GAKLTFKDLEGIGNALVKDQPAIPIYSN 214
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      201  NMTSDQIVALTFEESIKALKNVKSDDAIKGTGTSLTYNRLKDIVASELKQDSKLAVPYFK 260
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Qy      215  AKEIKNPPATVYDQAQTKMAHLDDWDSWPQDVPVTGYVSNMGYQLVIAWMGIPNSPTG 274
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      261  ADTIINPPAFNTVYDQAQTMKKEIDVWDSWPQDAKSGVSNWNGYQLVISMAGAPNK--N 318
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Qy      275  DNHTYILVLYNKYGDNDPFSHBNAGSTECTGKFNMYPOEWSGATLUNDQCTYQI 324

```

```

Db 678 MTNOGDWIDKSSSES-----LVHVGQDQNSAKLPNE-DFNVDDYAVS-----GYGLKPH--TY 726
Qy 682 PKIDQDPQSGQNTKNTKPNCG-----DKPAGKATP-----DNTNIDPSAOPSGON 728
Db 727 PTVDSGTCVSEAHGVLTVTKDGRKADKAPETPVSPTEGNSHVDKTKNPKCTSKPADNN 786
Qy 729 TNIDPSAOSQONTKNTVP-----GNEKQGNTPDAKOLPOTGNKSG 769
Db 787 ---OPSADKEDKPTNPTNPDPSPARTPPFYGDHSDNNDSSNDHHVAVPVKPTGDSVG 841

RESULT 3
SACB_BACAM STANDARD; PRT; 472 AA.
AC P21130;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Levansucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl
DE transferase) (Sucrose 6-fructosyl transferase).
CN SACB.
OS Bacillus amyloliquefaciens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1390;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 23844;
RX MEDLINE=91092506; PubMed=2265762;
RA Tang L.B., Lenstra R., Borchert T.V., Vasantha N.;
RT "Isolation and characterization of levansucrase-encoding gene from
RT Bacillus amyloliquefaciens.";
RL Gene 96:89-93(1990).
CC -1- CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL)(N) =
CC GLUCOSE + (2,6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN
CC ALSO ACT AS FRUCTOSYL ACCEPTORS).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- INDUCTION: BY SUCROSE.
CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; X52988; CAA37179.1; -
DR PIR; JQ0802; JQ0802.
DR PIR; S11739; S11739.
DR InterPro; IPR003469; Glyco_hydro.68.
DR Pfam; PF02435; Glyco_hydro.68.1.
KW Transferase; Glycosyltransferase; Signal.
FT SIGNAL 1 29
FT CHAIN 30 472 LEVANSUCRASE.
FT SEQUENCE 472 AA; 52859 MW; F38592D273677E7D CRC64;

Query Match 18.1%; Score 751.5; DB 1; Length 472;
Best Local Similarity 38.0%; Pred. No. 3e-33;
Matches 190; Conservative 88; Mismatches 171; Indels 51; Gaps 19;

Qy 144 NLKKNLPQAAKAVONAKIDAGSLTDQINELNKNFNSKAEGKALTFKDLGIGNAIVK 203
Db 2 NIKKIYQATVLTFTTALLAGGATQAFKENNQKAY-KETYGSHITRHDMLQIPRO-Q 58
Qy 204 QDPQAIYPSNAKEIKNMPATYTDQTKMAHLWDVDSHPVDPTVGYVSYNYNGYOLVI 263
Db 59 QNEKYQVPQDQSTIKN-----IESAKG-----LDVWDSWPLQN-ADGTVAEYNGYHWF 107
Qy 264 AMMGIPNSPTGDNIHLLYLLNKNYGDNDFSHWRNAGSIFGKTNVF-----QEWS 312

```

```

Db 108 ALAGSPKD-ADDTSIYFWYOKVGDNSIDSWKNAGRWF--KSDKFDANDPILKDOQEWS 164
Qy 313 GSAIVNDDGTIQLFFTSNDTSYKLDQRLATLNLNDDNDGVSISKSDVNYQVLFEGDG 372
Db 165 GSATFTSDCKIRLFYT--DYSGRHKYQSLTTAQVNVKSSDDTLKINGVEDHKTIFEDGDG 222
Qy 373 FHYQTYEQFAN-GKDRENDYCLRDHPHVOLENGDRDLYLFEANTGTED-YQSDDDQIYNWA 430
Db 223 KTYQNVQOFTIDEGNYTSGDNHTLRDPHYVE-DKGHKYLFVEANTGTENGQGESELFNKA 281
Qy 431 NYGGDDAFNKSSFKLNNKKDRELALANGALGILKLTNNQSKPKVEEYVSLVSTLMA 490
Db 282 YGGGTNFFRKESQKLOQSAKKRD-AELANGALGIIELNNDYT--LKKVWKPLITNTV 337
Qy 491 CDEV-----XXKLGDKYLYFSVTRVSRGSDRELTAKONTIVGDNVAMIGYVSDSLMGKYKP 546
Db 338 TDEIERANVFKMGKWLFT--DSRGSKMTI---DGINSNDIYMLGYVSNLSLTGPYKP 390
Qy 547 LNSGVVLFTASVPANMTATYIYAVPVAGHPDQVLITSTMSKNKDFASGGGNTATWAPSF 606
Db 391 LNKTLGLVQLMGLDNDVFTTYSHFAPQA-KGNVVTISYTNRGEF--EDKKATFGPSE 447
Qy 607 LVQINPDPTTTLARATNOG 626
Db 448 LMNIKGNKTSVVRNLSILEQG 467

RESULT 4
SACB_BACSU STANDARD; PRT; 473 AA.
AC P05655; P70984;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Levansucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl
DE transferase) (Sucrose 6-fructosyl transferase).
CN SACB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=85295507; PubMed=2993818;
RA Steinmetz M., Le Coq D., Aymerich S., Gonzy-Trebol G., Gay P.;
RT "The DNA sequence of the gene for the secreted Bacillus subtilis
RT enzyme levansucrase and its genetic control sites.";
RL Mol. Gen. Genet. 200:220-228(1985).
RN [2]
RP SEQUENCE FROM N.A.
RA Denizot F.C.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-62 FROM N.A.
RC STRAIN=168 / PY79;
RX MEDLINE=84178454; PubMed=6424671;
RA Fouet A., Arnaud M., Klier A., Rapoport G.;
RT "Characterization of the precursor form of the exocellular
RT levansucrase from Bacillus subtilis.";
RL Biochem. Biophys. Res. Commun. 119:795-800(1984).
RN [4]
RP SEQUENCE OF 1-68 FROM N.A.
RX MEDLINE=87008406; PubMed=2428811;
RA Shimotsu H., Henner D.J.;
RT "Modulation of Bacillus subtilis levansucrase gene expression by
RT sucrose and regulation of the steady-state mRNA level by sacU and
RT sacQ genes.";
RL J. Bacteriol. 168:380-388(1986).
CC -1- CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL)(N) =
CC GLUCOSE + (2,6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN
CC ALSO ACT AS FRUCTOSYL ACCEPTORS).
CC

```



CC -1- SUBCELLULAR LOCATION: Secreted  
 CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M14202; AAA22725.1; -  
 DR EMBL; Z94043; CAB08015.1; -  
 DR EMBL; K01987; AAA22724.1; -  
 DR EMBL; X02730; CAA26513.1; -  
 DR EMBL; Z99121; CAB15450.1; -  
 DR PIR; A25040; A25040; -  
 DR PIR; S07309; S07309; -  
 DR Subtilist; BG10388; sacB.  
 DR InterPro; IPR003469; Glyco\_hydro.68.  
 DR Pfam; PF02435; Glyco\_hydro.68; 1.  
 KW Transferase; Glycosyltransferase; Signal; Complete proteome.  
 FT SIGNAL 1 29  
 FT CHAIN 30 473 LEVANSUCRASE.  
 FT CONFLICT 12 12 V -> I (IN REF. 3).  
 SQ SEQUENCE 473 AA; 52971 MW; 3PBF2F571B41D5B0 CRC64;

Query Match 18.08; Score 746.5; DB 1; Length 473;  
 Best Local Similarity 36.3%; Pred. No. 5.5e-33;  
 Matches 185; Conservative 92; Mismatches 162; Indels 71; Gaps 18;

QY 144 NLKLNPOAAKAVQNAKIDAGSLDDQINELNKNFNSAEKAGLTKFDLEGIGNAIVK 203  
 DB 2 NIKKFAQATVLTFTTALLAGGATQAFKTN-----QKPKETYGISH-ITR 48  
 QY 204 QD-----POYAIPIYNAKEIKNNPATYTVDAQTGKMAHLVDWDSWPQDPVTGYV 253  
 DB 49 HDMLQIPQOKNEKYQVPEFDSSTIKNISSA-----KGLDWDWSPQLN-ADGTV 97  
 QY 254 SNYMGYQLVIAAMGIPNSPTGDNHLYLYNKYGNDFSHWRNAGSIFGTKETNVF----- 308  
 DB 98 ANYGHYHVFALAGDPKN-ADDTSIYFYQKVGSTDSWKNAGRVF--KDSKDFDANDS 154  
 QY 309 -----QEWSSAIVNDGTTQLFTSNDTSYKLNQDRLATATLNLVNDNGVSIKSDV 362  
 DB 155 ILKDOTQEWSSAIVNDGTTQLFTSNDTSYKLNQDRLATATLNLVNDNGVSIKSDV 362  
 QY 363 NYQVLFEGDGHYQTYEQFAN-GKDRNDYCLRDHPVQLENGDRYLVFEANTGTED-Y 420  
 DB 213 DYKSIFDGDGHYQTYEQFAN-GKDRNDYCLRDHPVQLENGDRYLVFEANTGTEDY 271  
 QY 421 QSDDOIYNWANGDGDAPNTKSSPKLLNNKDKRELALGALGILKLTNNQSKPKVEEV 480  
 DB 272 QGESLENKAYGKSTFFRQESQKLLQSDKNR-TAELANGALGIELNDYD---LKKV 327  
 QY 481 YSPLVSTLMACDEV-----XXKLGDKYYLFSVTRVSRGSDRELAKDNTIVGDNVAMIGYV 536  
 DB 328 MKPLIASNTVTDETERANVFMKNGKWLFTDSRGSKMTIDGTSND-----IYMLGYV 380  
 QY 537 SDSLMGKYKPLNNGGVLTASVPANWTATYVAVPVAGHPDOVLITSYMSKNKDFASGE 596  
 DB 381 SNSUTGYPKPLNTKGLVKMDLPNDVFTYSHFAVPOA-KGNVNVITSYMTNRGFIADK 439  
 QY 597 GNYATWAPSELVQINPDPTTTLARATNOQ 626  
 DB 440 --QSTRAPEFLNKGKTSVVKDSILEQ 467  
 RESULT 5  
 SABC\_BACST STANDARD; PRT; 473 AA.  
 ID SABC\_BACST  
 AC P94468;

DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Levansucrase precursor (Pc 2.4.1.10) (Beta-D-fructofuranosyl  
 DE transferase) (Sucrose 6-fructosyl transferase).  
 GN SABC OR SUBB.  
 OS Bacillus stearothermophilus.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Geobacillus.  
 OX NCBI\_TaxID=1472;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 12980;  
 RX MEDLINE=98007870; PubMed=9349714;  
 RA Li Y., Triccas J.A., Ferenci T.;  
 RT "A novel levansucrase-levanase gene cluster in Bacillus  
 RT stearothermophilus ATCC12980".  
 RL Biochim. Biophys. Acta 1353:203-208(1997).  
 CC -1- CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL)(N) -  
 CC GLUCOSE + (2,6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN  
 CC ALSO ACT AS FRUCTOSYL ACCEPTORS).  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U34874; AAB97111.1; -  
 DR InterPro; IPR003469; Glyco\_hydro.68.  
 KW Pfam; PF02435; Glyco\_hydro.68; 1.  
 KW Transferase; Glycosyltransferase; Signal.  
 FT SIGNAL 1 29 BY SIMILARITY.  
 FT CHAIN 30 473 LEVANSUCRASE.  
 SQ SEQUENCE 473 AA; 52854 MW; 12F7A272EC2B71E CRC64;

Query Match 17.6%; Score 730.5; DB 1; Length 473;  
 Best Local Similarity 35.9%; Pred. No. 3.9e-32;  
 Matches 183; Conservative 93; Mismatches 163; Indels 71; Gaps 19;

QY 144 NLKLNPOAAKAVQNAKIDAGSLDDQINELNKNFNSAEKAGLTKFDLEGIGNAIVK 203  
 DB 2 NIKKFAQATVLTFTTALLAGGATQAFKTN-----QKPKETYGISH-ITR 48  
 QY 204 QD-----POYAIPIYNAKEIKNNPATYTVDAQTGKMAHLVDWDSWPQDPVTGYV 253  
 DB 49 HDMLQIPQOKNEKYQVPEFDSSTIKNISSA-----KGLDWDWSPQLN-ADGTV 97  
 QY 254 SNYMGYQLVIAAMGIPNSPTGDNHLYLYNKYGNDFSHWRNAGSIFGTKETNVF----- 308  
 DB 98 ANYGHYHVFALAGDPKN-ADDTSIYFYQKVGSTDSWKTGPRVF--KDSKDFDANDS 154  
 QY 309 -----QEWSSAIVNDGTTQLFTSNDTSYKLNQDRLATATLNLVNDNGVSIKSDV 362  
 DB 155 ILKDOTQEWSSAIVNDGTTQLFTSNDTSYKLNQDRLATATLNLVNDNGVSIKSDV 362  
 QY 363 NYQVLFEGDGHYQTYEQFAN-GKDRNDYCLRDHPVQLENGDRYLVFEANTGTED-Y 420  
 DB 213 DYKSIFDGDGHYQTYEQFAN-GKDRNDYCLRDHPVQLENGDRYLVFEANTGTEDY 271  
 QY 421 QSDDOIYNWANGDGDAPNTKSSPKLLNNKDKRELALGALGILKLTNNQSKPKVEEV 480  
 DB 272 QGESLENKAYGKSTFFRQESQKLLQSDKNR-TAELANGALGIELNDYD---LKKV 327  
 QY 481 YSPLVSTLMACDEV-----XXKLGDKYYLFSVTRVSRGSDRELAKDNTIVGDNVAMIGYV 536  
 DB 328 MKPLIASNTVTDETERANVFMKNGKWLFTDSRGSKMTIDGTSND-----DGITSNDIYMLGYV 380

QY 537 SDSLKGYKPLNNSGVLTASYPANWRRTATSYAYVPVAGHDVOLTYSMSNKKDFASGE 596  
 Db 381 SSSLGPKYKPLNLTGLKMLDNDPNDVTYTSHEFAVPA-TGNVYVITSYMTNRGFIADK 439  
 QY 597 GNYATWAPSFVQINPDOTTTVLARATNOG 626  
 Db 440 --QSTFAPSFLLNIOGKTSVVKASILDG 467

Query Match 7.4%; Score 309; DB 1; Length 413;  
 Best Local Similarity 26.5%; Pred. No. 9e-10;  
 Matches 105; Conservative 71; Mismatches 150; Indels 70; Gaps 20;

RESULT 6  
 INVB\_ZYMMO STANDARD; PRT; 413 AA.  
 AC Q60115: Q60117; Q60125;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Extracellular sucrose (EC 3.2.1.26) (Beta-fructofuranosidase)  
 DE (Invertase) (Saccharase) (Protein B46).  
 GN SACC OR INVB OR SUC3.  
 OS Zymomonas mobilis.  
 OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;  
 OC Zymomonas.  
 ON NCBI\_TaxID=542;  
 RX [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN=ATCC 10988 / ZM1;  
 RX MEDLINE=94368848; PubMed=8086457;  
 RA Song K.B., Lee S.K., Joo H.K., Rhee S.-K.;  
 RT "Nucleotide and derived amino acid sequences of an extracellular  
 RT sucrose gene (invB) of Zymomonas mobilis ZM1 (ATCC10988).";  
 RL Biochim. Biophys. Acta 1219:163-166(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IFO 13756;  
 RX MEDLINE=95218269; PubMed=7766026;  
 RA Kyono K., Yanase H., Tonomura K., Kawasaki H., Sakai T.;  
 RT "Cloning and characterization of Zymomonas mobilis genes encoding  
 RT extracellular levansucrase and invertase.";  
 RL Biosci. Biotechnol. Biochem. 59:289-293(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN=NRRL B806;  
 RX MEDLINE=95297907; PubMed=7778976;  
 RA Kannan R., Mukundan G., Alt-Abdelkader N., Augier-Magro V.,  
 RA Baratti J., Gunasekaran P.;  
 RT "Molecular cloning and characterization of the extracellular sucrose  
 RT gene (sacc) of Zymomonas mobilis.";  
 RL Arch. Microbiol. 163:195-204(1995).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 31821 / ZM4 / CP4;  
 RA Ahn J.Y., Kang H.S.;  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-  
 CC fructofuranoside residues in beta-D-fructofuranosides.  
 CC -!- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF081588; AAA61488.1; -;  
 DR EMBL; D17524; BAA04476.1; -;  
 DR EMBL; L33403; AAC36942.1; -;  
 DR EMBL; AF313764; AAG29871.1; -;  
 DR InterPro; IPR003469; Glyco\_hydro\_68.  
 DR Pfam; PF02435; Glyco\_hydro\_68; 1.  
 KW Hydrolase; Glycosidase.  
 FT CONFLICT 85 G -> A (IN REF. 3).

FT CONFLICT 105 105 S -> L (IN REF. 2).  
 FT CONFLICT 108 108 T -> S (IN REF. 2).  
 FT CONFLICT 404 407 PVWP -> LGMA (IN REF. 3).  
 SO SEQUENCE 413 AA; 46100 MW; BACC12D167332B47 CRC64;

Query Match 7.4%; Score 309; DB 1; Length 413;  
 Best Local Similarity 26.5%; Pred. No. 9e-10;  
 Matches 105; Conservative 71; Mismatches 150; Indels 70; Gaps 20;

QY 239 VWDSPVQDPVTGVSNMCGYQVLVIAMGIPNSPTGDNHLYLLNKKYGDNDGSHWRNAGS 298  
 Db 42 LWDTPWLRD-INGNPVSFKGNVIFSLVADRNPWDRSHARIGYFYSKDGKSWYGGH 100  
 QY 299 IFGTKETNVQEWSSAIV--NDDGTIQLFTTSNDTSDYKLNDRLATATL---NLNVDD 353  
 Db 101 LLOESANTRTAEMSGGTIMAPGSRNQVETFTS---TLFDKNGVREAVAATKGRIVADS 157  
 QY 354 NGVSIKSDNYQVLFEQDGFHYQTYEFGANGKRENDYDCLRDPHV-VQLENGDRYLVFE 412  
 Db 158 EGVWFKGFQDOSTDIFQADGLFYQNY-----AENNLWNRDPHFVFNPEDEGTYALFE 209  
 QY 413 ANTGTEDYQSDDOIYNWANYGDDDAFNKSKSEKLLNKKKRELKAGLALGILKLTNNQ 472  
 Db 210 ANYAT--VRGEDI-----GEDEIGVPANTVV---PKD---ANLCSASIGARCL--- 252  
 QY 473 SKPKVE-EVYSPLVSLTMACDEVXXK---LGDKYILFVSRTVSRGSDRELTAKDNTIVG 527  
 Db 253 SPORTEWELLPLLTAFGVNDQMERPHVIFQNGLTLYFTISHDSTYAD-GLTGS----- 306  
 QY 528 DNYAMIGYVSDS-LMGKYKPLNNSGVLTASYPANWRRTATSYAYVPVAGHDVOLTYSY 586  
 Db 307 ---GLYGFVSENGIFGYPEPLNGSLVGG--PASQPTAYAHYIM-----NNGLVSEF 355  
 QY 587 M-----SNKDFASGEGNTATWAPSFVQINPDOT 615  
 Db 356 INELIDPKSKVYAGG-----SLAPTVRVELOGHET 386

RESULT 7  
 SACC\_ZYMMO STANDARD; PRT; 423 AA.  
 AC Q60114: Q60116; Q60487;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Levansucrase (EC 2.4.1.10) (Beta-D-fructofuranosyl transferase)  
 DE (Sucrose 6-fructosyl transferase).  
 GN SACC OR LEVU OR SUC2.  
 OS Zymomonas mobilis.  
 OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;  
 OC Zymomonas.  
 ON NCBI\_TaxID=542;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 10988 / ZM1;  
 RX MEDLINE=93305726; PubMed=8318541;  
 RA Song K.B., Joo H.K., Rhee S.-K.;  
 RT "Nucleotide sequence of levansucrase gene (levU) of Zymomonas mobilis  
 RT ZM1 (ATCC10988).";  
 RL Biochim. Biophys. Acta 1173:320-324(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN=IFO 13756;  
 RX MEDLINE=95218269; PubMed=7766026;  
 RA Kyono K., Yanase H., Tonomura K., Kawasaki H., Sakai T.;  
 RT "Cloning and characterization of Zymomonas mobilis genes encoding  
 RT extracellular levansucrase and invertase.";  
 RL Biosci. Biotechnol. Biochem. 59:289-293(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL B806;  
 RX MEDLINE=95297907; PubMed=7778976;  
 RA Kannan R., Mukundan G., Alt-Abdelkader N., Augier-Magro V.,  
 RA Baratti J., Gunasekaran P.;  
 RT "Molecular cloning and characterization of the extracellular sucrose  
 RT gene (sacc) of Zymomonas mobilis.";  
 RL Arch. Microbiol. 163:195-204(1995).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 31821 / ZM4 / CP4;  
 RA Ahn J.Y., Kang H.S.;  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-  
 CC fructofuranoside residues in beta-D-fructofuranosides.  
 CC -!- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF081588; AAA61488.1; -;  
 DR EMBL; D17524; BAA04476.1; -;  
 DR EMBL; L33403; AAC36942.1; -;  
 DR EMBL; AF313764; AAG29871.1; -;  
 DR InterPro; IPR003469; Glyco\_hydro\_68.  
 DR Pfam; PF02435; Glyco\_hydro\_68; 1.  
 KW Hydrolase; Glycosidase.  
 FT CONFLICT 85 G -> A (IN REF. 3).

```

DR EMEL; C931408; AAC304361;
DR InterPro: IPRO03469; Glyco_hydro_68.
DR Pfam: PF02435; Glyco_hydro_68; 1.
DR K Transferase; Glycosyltransferase.
SQ SEQUENCE      415 AA; 45939 MW; E028828813D13A74 CRC64;

Query Match          5.98; Score 245; DB 1; Length 415;
Best local Similarity 23.8%; Pred. No. 2.3e-06;
Matches 100; Conservative 72; Mismatches 159; Indels 90; Gaps 20

Qy 219 KNMPATY--TVDAQTGKA-HLDVWDSPVPDPTGYGVSYNTMGVYOLVIAMGIPNSPTGD 275
   ||| | | | | | | | :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 21  ENDPTTQPIVDADFPVMSEVFITWTPILRS-LDGTVVSDGHSVITFLTAQRNN---N 76
   ||| | | | | | | | :||| :||| :||| :||| :||| :||| :||| :||| :|||

Qy 276 NHLYL-----LYNKYGDNDFSHRN-----AGSIFGKTETNVFEWSGSAI 316
   ||| | | | | | | | :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 77  NSEVLDAEGNYDITSWNRRHGRARICVYSRTGKDWITFGGRVMAEVSPTSREWAGTPI 136
   ||| | | | | | | | :||| :||| :||| :||| :||| :||| :||| :||| :|||

Qy 317 -VNDDGTTQLFTSNDSYKLNDORLATNLNVDNGVSIKSVDKNQVLFECGDGHY 375
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 137 LLNEDGGTDLYTCVTGP-----ATIAKVRGKVLTSEEGVTLAGFNEVKLSFDAGVYY 190
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

Qy 376 QTYEQFANGKRENDYCLRDPH-VVQLENGDRYLVEFANTGTEDYQSDQIQINANYNG 434
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 191 QT-----ESQPNWFNPDPSPFDHPDGKLVYFEGNVAGE-----RG 228
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

Qy 435 DDAFNISFKLLANKKDELALANGALGL----KLTNNSOKPVEEVYSLPVSTLMAC 491
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 229 SHVIGKQEMGTLPPCHRDVGVARQAGCIGMAVAKDLSGDE----WEILPLVITAVGN 283
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

Qy 492 DEVXX-----LGDXYLESVTRVSRGSDBELTAKONTIVGNVAMIGVSSLSMCKYKPL 547
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 284 DQTERPHEVFQDGKYVLYFTISHKFTYAD-TGTPD-----GVYGLSDNLTPGSPM 334
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```

QY 548 NNSGVLTASVPANWRKATYSYAVVPAGHPDQVLITSYMKNKDRASGRGNV---ATWAP 604
DB 335 NGSGLVL--GNPPSOPFQYISHCVMP-----NGLVTSFIDN--VPTSDGNRYRIGGTEAP 384

QY 605 S 605
DB 385 T 385

RESULT 9
SACB_PSESG STANDARD; PRT; 415 AA.
AC OS2408;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Levansucrase (EC 2.4.1.10) (Beta-D-fructofuranosyl transferase)
DE (Sucrose 6-fructosyl transferase).
GN LSC.
OS Pseudomonas syringae (pv. glyciniae).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=318;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=PG4180;
RX MEDLINE=98394981; PubMed=9726857;
RA Hettwer U., Jaekel F.R., Boch J., Meyer M., Rudolph K., Ullrich M.S.;
RT "Cloning, nucleotide sequence, and expression in Escherichia coli of
RT levansucrase genes from the plant pathogens Pseudomonas syringae pv.
RT glyciniae and P. syringae pv. phaseolicola.";
RL Appl. Environ. Microbiol. 64:3180-3187(1998).
CC -1- CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL)(N) -
CC GLUCOSE + (2,6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN
CC ALSO ACT AS FRUCTOSYL ACCEPTORS).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF037443; AAC36056.1;
CC InterPro: IPR003469; Glyco_hydro_68.
CC Pfam: PF02435; Glyco_hydro_68; 1.
CC Transferase; Glycosyltransferase.
CC SEQUENCE 415 AA; 4584 MW; 839B686AC80610CF CRC64;

Query Match 5.8%; Score 242.5; DB 1; Length 415;
Best Local Similarity 23.6%; Pred. No. 3.2e-06;
Matches 96; Conservative 68; Mismatches 154; Indels 89; Gaps 19;

QY 239 VWDSPVQDPVTGYVSNMGYQVLVAMG--IPNSPTGDHNIYLL----- 281
DB 44 IWDTHPLRE-LDGTVVVNGSVIYTLTADRHDPDQVVGANGRYDIKRDWEDRHGRAM 102

QY 282 ---YNKYGDNDFSHWRNAGSIFGTKETNVFQWGS--AIVNDGDTIQLFFTSNDTSDYKL 337
DB 103 CYWYSRTGKD---WIFGGRVMAEGVSPPTREWAGTVPVLLNDKGDIDLYITCVTPG---- 154

QY 338 NDQRLATATLNLVNDNGVSKSDVNYQVLFEGDGFHYQTFYQFANGKDRENDYCLRD 397
DB 155 ---AATAKVRGRIVTSDKGVELKDFTEVTLFEADGKYIYT-----EAQNSTWNRDP 204

QY 398 H-VVQLENGDYLVPFANTGTGTDYSDQIYNWANYGGDADFNIKSSFKLLNNKKDRELA 456
DB 205 SPFIDPNDGKLYMVEGNY-----AGERGTHVGAALGLGVPPPGHEET 247

```

```

QY 457 GLANGALGILKLTNNOSKPKVE-EVTSPLVSTLMACDEVXXK-----LGDKYILFSVTRVS 511
DB 248 GGARFQVCGIGLAVAKDLSGDEWEIILPPLVAVGVNDQTERPHYVFDQDKYILFTIS--- 304

QY 512 RGSDELTAKDNTIVGDNVAMIGYVSDSLMGKYKPLNNSGVLTASVPANWRKATYSYVA 571
DB 305 -----HKFTYADGVTPGDGV--YGFVGEHLFGPYRPMNASGLVL-GNPPAQ-PPQTYSHCV 356

QY 572 VPVAGHPDQVLITSYM-----SNKDFASGEGNYATWAPSLVQINPD 613
DB 357 MP-----NGLVTSFIDSPTSCEDYRIG-----GTEAPTVRILLEGD 393

RESULT 10
SACB_PSESH STANDARD; PRT; 431 AA.
AC O68609;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Levansucrase (EC 2.4.1.10) (Beta-D-fructofuranosyl transferase)
DE (Sucrose 6-fructosyl transferase).
GN LSC.
OS Pseudomonas syringae (pv. phaseolicola).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=319;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NCPPB 1321;
RX MEDLINE=98394981; PubMed=9726857;
RA Hettwer U., Jaekel F.R., Boch J., Meyer M., Rudolph K., Ullrich M.S.;
RT "Cloning, nucleotide sequence, and expression in Escherichia coli of
RT levansucrase genes from the plant pathogens Pseudomonas syringae pv.
RT glyciniae and P. syringae pv. phaseolicola.";
RL Appl. Environ. Microbiol. 64:3180-3187(1998).
CC -1- CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL)(N) -
CC GLUCOSE + (2,6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN
CC ALSO ACT AS FRUCTOSYL ACCEPTORS).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF052289; AAC36063.1;
CC InterPro: IPR003469; Glyco_hydro_68.
CC Pfam: PF02435; Glyco_hydro_68; 1.
CC Transferase; Glycosyltransferase.
CC SEQUENCE 431 AA; 47603 MW; 795FEA246D40C40F CRC64;

Query Match 5.8%; Score 240.5; DB 1; Length 431;
Best Local Similarity 23.3%; Pred. No. 4.2e-06;
Matches 95; Conservative 68; Mismatches 154; Indels 91; Gaps 18;

QY 239 VWDSPVQDPVTGYVSNMGYQVLVAMG--IPNSPTGDHNIYLLYN-----KY 285
DB 60 IWDTHPLRE-LDGTVVVNGSVIYTLTADRHDPDQ-----YLDANGRYDIKRDWEDRH 113

QY 286 GDNDFSHWRN-----AGSIFGTKETNVFQWGS--VNDGDTIQLFFTSNDTSDYK 336
DB 114 GRARMSYWSRTGKDWIFGGRVMAEGVSPPTREWAGTVPVLLNDKGDIDLYITCVTPG--- 170

QY 337 LNDQRLATATLNLVNDNGVSKSDVNYQVLFEGDGFHYQTFYQFANGKDRENDYCLRD 396
DB 171 ---AATAKVRGRIVTSDQVELKDFTEVTLFEADGTYIYT-----EAQNSSWNRD 219

```

```

QY 397 PH-VVLENGDRYLVPFANTGTEDYQSDDOIYNWANYGGDDAFNIKSSFKLLNKKDREL 455
Db 220 PPSFIDPNDGKLYMVEGNY-----AGERSGHTYGAALGFPVPEGHED 262
QY 456 AGLANGALGTLKTNOSKPVE-EVYSLPVLSTLMACDEVXK-----LGDKYLYFVTRV 510
Db 263 VGAROVCGICGLAVAKDLSGEWELPLVTVAVGVNDQTERPHYFQDGKYLFTISHK 322
QY 511 SRGSDRELTAQNTIYGVNDVAMIGYVSDSLMGKYKPLNNSGVVLTASVPANMRTATYSY 570
Db 323 FTVAE-GLTGPD-----GVYGFGEHLFGPYRPMNASGLVL--GNPPEQPFQYSHC 371
QY 571 ANPVAGHPQOVLITSYMSN-----KDFASGEGNYATWAPSLVQINPD 613
Db 372 VMP-----NGLVTSFIDSVPTEGEDYRIG-----GTEAPTVRILLKGD 409

RESULT 11
SACB_ERWAM STANDARD; PRT; 415 AA.
AC Q46654;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Levanucrase (EC 2.4.1.10) (Beta-D-fructofuranosyl transferase)
DE (Sucrose 6-fructosyl transferase).
GN LSC.
OS Erwinia amylovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Erwinia.
OX NCBI_TaxID=552;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EA774;
RA Geier G., Geider K.K.;
RT "Characterization and influence on virulence of the levansucrase gene
from the fireblight pathogen Erwinia amylovora.";
RL Physiol. Mol. Plant Pathol. 42:387-404(1993).
CC -1- CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL)(N) =
GLUCOSE + (2,6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN
ALSO ACT AS FRUCTOSYL ACCEPTORS).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; X75079; CAAS2972.1;
DR InterPro; IPR003469; Glyco_hydro_68.
DR Pfam; PF02435; Glyco_hydro_68; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 415 AA; 46408 MW; 4FF564F6E0607FEB CRC64;

Query Match 5.7%; Score 236.5; DB 1; Length 415;
Best Local Similarity 23.4%; Pred. No. 6.6e-06;
Matches 96; Conservative 67; Mismatches 152; Indels 95; Gaps 20;

QY 239 WDSWFPQDPVTGVSYNMGVOLVIAMMGPNSPT-----GDNHI 278
Db 44 IWDTPMLRD-FDGEIISVNGWCIIITLADRNTDNPQFQENGWYDTRWDHRGRARI 102
QY 279 YLLNRYGDNDFSHWRNAGSFGTKETNVQWMSGSAI-VNDQGTIQIQLFTSNDTSYKL 337
Db 103 CYWTSRTGKD----WIFGGRVNAEGVAPTTREWAGTPLLNDRGDIDLYTCVTPG---- 154
QY 338 NDQRLATATLNLVNDNGSVKTSVDNYQVLFEGDGFHYQYVEQFANGKDRNDYCLDRDP 397

```

```

Db 155 --ATIAKVRGKIVTSDQSVSLBGFQQVTSLSADSGTIYQTEBQ-----NAFWNRDP 204
QY 398 H-VVLENGDRYLVPFANT-----GTEDYQSDDOIYNWANYGGDDAFNIKSSFKLLNKKD 452
Db 205 SPFIDRNDGKLYMVEGNYAGPRGSHEITQ-----AEMG-----NVPPEGYEDVGAKY 252
QY 453 RELAGLANGALGTLKTNOSKPVEEYSLPVLSTLMACDEVXK-----LGDKYLYFVST 508
Db 253 QA-----GCVG-LAVAKDLSGSE-WQILPLPLITAVGVNDQTERPHYFQDGKYLFTIS 304
QY 509 RVSRGSDRELTAQNTIYGVNDVAMIGYVSDSLMGKYKPLNNSGVVLTASVPANMRTATYS 568
Db 305 -----HKYTEADNLTPDGV--YGFVSDKLTGPTPMNSSLVL--GNPSPQFQYTS 353
QY 569 YVAVPVAGHPDQVLITSYMSN-----KDFASGEGNYATWAPSLVQINPD 613
Db 354 HYVMP-----NGLVTSFIDSVPWPKGDYRIG-----GTEAPTVKILLKGD 393

RESULT 12
SACB_ACEDI STANDARD; PRT; 584 AA.
AC Q43998;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Levanucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl
transferase) (Sucrose 6-fructosyl transferase).
GN LSDA.
OS Acetobacter diazotrophicus (Gluconacetobacter diazotrophicus).
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
OC Gluconacetobacter.
OX NCBI_TaxID=33996;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 52-61.
RC STRAIN=SRT4 / CBS 550.94;
RX MEDLINE=96253999; PubMed=8704949;
RA Arieta J., Hernandez L., Coego A., Suarez V., Balmori E.,
Menendez C., Petit-Glatron M.-F., Chamber R., Selman-Housein G.;
RT "Molecular characterization of the levansucrase gene from the
endophytic sugarcane bacterium Acetobacter diazotrophicus SRT4.";
RL Microbiology 142:1077-1085(1996).
CC -1- FUNCTION: RELEASES FRUCTOOLIGOSACCHARIDES AND LEVAN, A HIGH-
MOLECULAR-MASS FRUCTOSYL POLYMER, FROM SUCROSE. IT ACTS MORE AS A
SUCROSE HYDROLASE THAN AS A FRUCTAN POLYMERASE.
CC -1- CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL)(N) =
GLUCOSE + (2,6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN
ALSO ACT AS FRUCTOSYL ACCEPTORS).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- PTM: THE N-TERMINAL IS BLOCKED.
CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; L41732; AAB36606.1;
DR InterPro; IPR003469; Glyco_hydro_68.
DR Pfam; PF02435; Glyco_hydro_68; 1.
KW Transferase; Glycosyltransferase; Signal.
FT SIGNAL 1 30 POTENTIAL.
FT PROPEP 31 51
FT CHAIN 52 584 LEVANSUCRASE.
SQ SEQUENCE 584 AA; 63525 MW; CBFBF4139AD0B8CE CRC64;

Query Match 4.7%; Score 197; DB 1; Length 584;
Best Local Similarity 23.8%; Pred. No. 0.0013;
Matches 117; Conservative 78; Mismatches 175; Indels 122; Gaps 31;

```

```
QY 220 NMPATYVDAQTGKMAHLD--VMSWVPQDPVTCYVSNYMGYQLVIAIMGIPNSPTG--D 275
DB 119 NIPADFV-----INPDVWVDWTLDKHAQFS-YNGWEVIECLTADPNAGYFDD 170
QY 276 NHY-----LYNK-----YGDNFSHRNAGSFGTKETNVFQWSSGSAI 316
DB 171 RHVHARIGFYRAGIPASRRPVNGWYTGCHLFPDGCASQVAGQYTNQ-AEWSGSSR 229
QY 317 VND-DGTIOLEF-----SNDTSDYKLN-DQRLATATL-NLNVDDNGVSIKSDVNDYQVLF 368
DB 230 LMQIHGNTSVFTDVAFNDAANNITPPQAITQTLGRHADFNHWFTGTAHTPILL 289
QY 369 EGDGFHYQTYEQFANGKRDENDYCLDRPHVQ--LENGDRYLVFEANT-----GTEDYQS 422
DB 290 QPDGVLYQ-----NG--AQNEEFNFRDPTFEDPKHPGVNMYVFEGTAGQRCVANCTE 341
QY 423 DQQTYNWANGGDDAFNIKSFLLNNKKDRELA--GLANGALGILKLTNNQSKPKVEEV 480
DB 342 AD-----LGRPNDDP-NAETLQELVDSGAYYQKANIOLA-----IATDSTLSKWK---F 386
QY 481 YSPLVSTLMACDEVXXKL-----GDKYYLFSV---TRVSRGSDRELTAKDNTIVGDNVAMI 533
DB 387 LSPILSANCVDQTERQVYLHNGKYYIFITISHETTAAGVD-----GPD-GVY 434
QY 534 GYVSDSLMGYKPLNNSGVVLTASVPANWRTA-----TYSYAVPVAG 576
DB 435 GFVGDGIRSDFPQNP-NYGSGLTMGNPTDLNTAAGTDEDPSPQNPRAFQSYSHYVMPGG- 492
QY 577 HPDQVLTYSNKNKDPASGEGNATWAPSLVQINPDPTTVTLARNTNOCGWVDD--SS 634
DB 493 -----LVESFIDTYENRRG-----GTLAPTVRVRI-AQNASAVDLRYGNGGLGYGDIPAN 542
QY 635 RND-NMLGVLKE 645
DB 543 RADVNIAGFIQD 554

RESULT 13
ID KYK1_DICDI STANDARD; PRT; 1584 AA.
AC P18160; 1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Non-receptor tyrosine kinase spore lysis A (EC 2.7.1.112) (Tyrosine-
DE protein kinase 1).
GN PYKA OR SPLA OR DPYK1.
OS Dictyostelium discoideum (Slime mold). Dictyostelium.
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JH10.
RA MEDLINE=97053827; PubMed=8898241;
RX Nuckolls G.H., Osheroov N., Loomis W.F., Spudich J.A.;
RT "The Dictyostelium dual-specificity kinase splA is essential for
RT spore differentiation.";
RL Development 122:3295-3305(1996).
RN [2]
RP SEQUENCE OF 1248-1584 FROM N.A.
RX MEDLINE=90287147; PubMed=1972546;
RA Tan J.L., Spudich J.A.;
RT "Developmentally regulated protein-tyrosine kinase genes in
RT Dictyostelium discoideum.";
RL Mol. Cell. Biol. 10:3578-3583(1990).
CC -1- FUNCTION: ESSENTIAL FOR SPORE DIFFERENTIATION.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT DEVELOPMENT WITH A PEAK
CC DURING THE MOUND STAGE OF MORPHOGENESIS.
CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES BUT ALSO TO
CC SERINE/THREONINE PROTEIN KINASES.
```

```
CC -1- SIMILARITY: CONTAINS 1 SAM DOMAIN.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U32174; AAB41125.1;
CC PIR: M33785; AAA33202.1;
CC HSP; P11362; IAGW.
CC DictyDb; DD03010; PYKA.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR001660; SAM.
CC InterPro; IPR003877; SPRY.
CC InterPro; IPR003878; SPRY_domain.
CC InterPro; IPR004040; STY_pkinase.
CC InterPro; IPR001245; TYR_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF00536; SAM; 1.
CC Pfam; PF00622; SPRY; 3.
CC SMART; SM00454; SAM; 1.
CC SMART; SM00449; SPRY; 3.
CC SMART; SM00221; STYKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE; PS00105; SAM_DOMAIN; 1.
CC Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
KW DOMAIN 908 972 SAM.
FT DOMAIN 403 420 POLY-ASN.
FT DOMAIN 428 435 POLY-THR.
FT DOMAIN 449 480 POLY-ASN.
FT DOMAIN 483 491 POLY-ASN.
FT DOMAIN 494 508 POLY-ASN.
FT DOMAIN 512 532 POLY-ASN.
FT DOMAIN 596 600 POLY-ASN.
FT DOMAIN 808 811 POLY-PHE.
FT DOMAIN 1026 1029 POLY-SER.
FT DOMAIN 1195 1210 POLY-ASN.
FT DOMAIN 1215 1220 POLY-GLN.
FT DOMAIN 1224 1233 POLY-GLN.
FT DOMAIN 1266 1274 POLY-PRO.
FT DOMAIN 1289 1561 PROTEIN_KINASE.
FT NP_BIND 1295 1303 ATP (BY SIMILARITY).
FT BINDING 1316 1316 ATP (BY SIMILARITY).
FT ACT_SITE 1417 1417 BY SIMILARITY.
FT CONFLICT 1248 1248 D -> R (IN REF. 2).
FT CONFLICT 1435 1435 V -> L (IN REF. 2).
SQ SEQUENCE 1584 AA; 174304 MW; 5D1589458D8E01E3 CRC64;

Query Match 4.6%; Score 190.5; DB 1; Length 1584;
Best Local Similarity 19.3%; Pred. No. 0.011;
Matches 165; Conservative 103; Mismatches 283; Indels 303; Gaps 40;

QY 15 SASILMGVVYTAHADQVESNNYNGVAEVNTERQANQIGVDGKIISANSNTTSGST--N 71
DB 396 STSIL-----VPGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 443
QY 72 QESSATNTNTENAVVNESKNTNTENAVVNEKNKNTNTENAVVNEKNKNTNTENDNSQLK 131
DB 444 EDISSNNNNNN--NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 501
QY 132 TNEQPSATQALKKLNQAKAVON-----AKIDAGSLTDDQINEL--NKINESK 181
DB 502 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 561
QY 182 SAEKGAKLTKDLEGIGN--AIVKQ-DPOVAIPYSNAKEIKKNMPATYTVDAQTGKMAHLD 238
```

Db 562 NSSSGS-----IGNNSSIIKQRPSPHSI-----NGP-----587  
 QY 239 VWDSPVQDPVTVGYVSNMGYQLVITAMGIPNSPTGDNHIIYLLNKYGDNDPSHWRNAGS 298  
 Db 588 -----LMLPPSSTNNNNIYSYNS-----TTAGS 612  
 QY 299 -----IFGKTETNVQEWGSAIYVNDGTTQLFTTSNDTSYKLNQDLATLNLVDDN- 354  
 Db 613 STTILPTLNRPIF---GNTTSNNSSSTLSVGGNNLLGRHCQSLPITASTNHTLSSSL 668  
 QY 355 GVSYSKS-----VDNQVLFEGGFHYQVTFQFANGKDRNDYDCLRDHPVVOL 402  
 Db 669 GVSFSPPSSPSTPKRPIVNSSEDL---GF-VQTFQD---QDGPSPAWRCGKSIXT 719  
 QY 403 ENDRYLVEFANTGTEDYQSDQIYNWANYGGDDAFNIKSPKLL-----NNKKDRE 454  
 Db 720 KODITLTIKKTSVA-----MADRPSSNSSSTICYFEVYLEGHDKKSI 765  
 QY 455 LAGLANGALGILKLTNNQSK-----PKVEEYVSPV-----STLMACD 492  
 Db 766 TVGLSHSTYPTFKHIGREPKSYGFSSEGEKYGGSEICEPYGPFPPFDGDSIASSCVIGCG 825  
 QY 493 EVXXKLGDKYVLFVTRVSRGSDRELTAQNTIVGDNVAMIGYVSDSLMGKYKPLNNSGV 552  
 Db 826 -INTSTRDIFF-----TNGHYLG--VAFSRVTSPL---YPSISFRGV 863  
 QY 553 V-----LTASVPANMRTATYSYVAVPVAGHPDQVLITSYMSKNKDEA 593  
 Db 864 VGLCVATPFGGHFRFNIEDLPGISPVWTEAL-----GPR-----900  
 QY 594 SGBGNATWAPSLVQINPDITTLARATNOGDWYDSSRDNDMLGVLKEGAANSAL 653  
 Db 901 QGSG-FKNMAP-----NOVAIWLESFNYGQ---YRKNFRDNISGRHLEITH-AML 947  
 QY 654 PGWGWKPVWDLNRLSPGLKHPQVQPKIDQ-----PDQPSGQNTKNVTP 701  
 Db 948 KND-----LGIEFYGHREDIINRLRMQIWNDKSPDSYP-----KIAI 986  
 QY 702 GNGDK---PAGKATPDNTNI-----DPSAQPSGQNTNDP-----SAQXSGQ 740  
 Db 987 DSSDKIRWPAAGSGSGGINISGGVWIGSSGSDGTEITSSSKNIRPKYSYQKEIEDR 1046  
 QY 741 NTKNVTPEGNEKCK 754  
 Db 1047 NRRSTISGGEKKNK 1060

RESULT 14  
 AAC2\_DICDI STANDARD; PRT; 448 AA.  
 AC P14196;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE AAC-rich mRNA clone AAC11 protein (Fragment).  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
 OX NCBI\_TaxID=44689;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90066348; PubMed=2511421;  
 RA Shaw D.R., Richter H., Glorda R., Ohmachi T., Ennis H.L.;  
 RT "Nucleotide sequences of Dictyostelium discoideum developmentally  
 regulated cDNAs rich in (AAC) imply proteins that contain clusters of  
 asparagine, glutamine, or threonine.";  
 RL Mol. Gen. Genet. 218:453-459(1989).  
 CC -1- DEVELOPMENTAL STAGE: THE CONCENTRATION OF AAC-RICH MRNAS IS LOW  
 CC IN DORMANT SPORES AND GROWING CELLS BUT INCREASES DURING  
 CC SPORE-GERMINATION AND MULTICELLULAR DEVELOPMENT.  
 CC -1- MISCELLANEOUS: SEVERAL PROTEINS DERIVE FROM AAC-RICH MRNA, WHICH,  
 CC DUE TO A FRAMESHIFT ALSO HAVE ACA AND CAA CODONS AND THUS ARE  
 CC ASN-, THR- OR GLN-RICH.  
 CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; X16522; CAA34529.1; -.  
 DR PIR; S05355; S05355.  
 DR DictyDb; DD05006; -.  
 DR InterPro; IPR000637; AT\_hook.  
 DR InterPro; IPR000116; Highmobility\_IV.  
 DR Pfam; PF02178; AT\_hook; 4.  
 DR ProDom; PD005593; Highmobility\_IV; 1.  
 DR SMART; SM00384; AT\_hook; 4.  
 KW Repeat.  
 FT NON\_TER 1 1 GLN-RICH.  
 FT DOMAIN 5 25 ASN-RICH.  
 FT DOMAIN 115 144 ASN-RICH.  
 FT DOMAIN 268 384 ASN-RICH.  
 SQ SEQUENCE 448 AA; 48636 MW; D8ED80D910D99817 CRC64;

Query Match 4.6%; Score 190; DB 1; Length 448;  
 Best Local Similarity 25.3%; Pred. No. 0.0022;  
 Matches 56; Conservative 30; Mismatches 85; Indels 50; Gaps 7;  
 QY 19 LMGGVVTAAHQVSNNGVAVENRTERQAGIQGVGDKIISANSNTSGTQESSATN 78  
 Db 262 LGNGILSNNN 312  
 QY 79 NTEAVVNESKNTNTENAVVNEKNTNTENAVVNEKNTNTENDNSQLKLTNNQPS 138  
 Db 313 NTNNNNNTNN 367  
 QY 139 AATQANLKLNPQAAKAVQNAKIDAGSLTDDQINELNKFNSKSAEKGAKLTFKDLGIG 198  
 Db 368 NNNHNNNNNTS-----NNSNTTATTAPGNNLT-NSLNAG 406  
 QY 199 N-----AIVQDDPOYAIYPYNAKEIKNP-ATYTVDAQ 230  
 Db 407 NLGNLGRVSGHSSDP-----NNPNAQKSPDSTNTMDFQ 441

RESULT 15  
 DEXT\_STRMU STANDARD; PRT; 850 AA.  
 AC Q54443;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Dextranase precursor (EC 3.2.1.11) (Alpha-L-6-glucan-6-  
 DE glucanohydrolase).  
 GN DEXA.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=INGBRITT;  
 RX MEDLINE=96245544; PubMed=8657012;  
 RA Igarashi T., Yamamoto A., Goto N.;  
 RT "Sequence analysis of the Streptococcus mutans Ingbritt dexa gene  
 RT encoding extracellular dextranase.";  
 RL Microbiol. Immunol. 39:853-860(1995).  
 CC [2]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN=INGBRITT;  
 CC MEDLINE=96081507; PubMed=8535521;  
 CC Colby S.M., Whiting G.C., Tao L., Russell R.R.B.;  
 CC "Insertional inactivation of the Streptococcus mutans dexa



RT (dextranase) gene results in altered adherence and dextran  
catabolism. 141:2929-2936(1995).  
RL Microbiology 141:2929-2936(1995).  
CC -1- FUNCTION: MAY PLAY A ROLE IN SUCROSE-INDEPENDENT ADHERENCE TO THE  
CC PELICLE-COATED TOOTH SURFACE.  
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,6-alpha-D-glucosidic  
CC linkages in dextran.  
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.  
CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS  
CC IN THE REGION OF THE MEMBRANE ANCHOR.  
CC -1- SIMILARITY: BELONGS TO FAMILY 66 OF GLYCOSYL HYDROLASES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL: D49430; BAA08409.1; .  
DR InterPro: IPR001899; Gram\_pos\_anchor.  
DR PROSITE: PS00343; GRAM\_POS\_ANCHORING; 1.  
KW Hydrolase; Glycosidase; Signal; Transmembrane.  
FT SIGNAL 1 ?  
FT CHAIN ? 850 DEXTRANASE.  
FT DOMAIN ? 823 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 824 844 POTENTIAL.  
FT DOMAIN 845 850 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 811 816 CONSERVED IN GRAM-POSITIVE COCCI SURFACE  
FT PROTEINS.  
SQ SEQUENCE 850 AA; 94538 MW; D653CAA39C93D2AC CRC64;

Query Match 4.5%; Score 185; DB 1; Length 850;  
Best Local Similarity 21.0%; Pred. No. 0.0094;  
Matches 201; Conservative 108; Mismatches 290; Indels 358; Gaps 55;  
Qy 26 AHADQVESNNYNGVAEVNTERQANGQIGVDGKIISANSNT-----TSGSTNQESS---A 76  
Db 40 AAAYOTES-----ASIDSNEQEQS-----VSANTNTPQAKKLSNNSHQEPQMVS 85  
Qy 77 TNNTENAVVNESKNTNN-----TENAV-----VNEKNKNTNTENAVVNENKNTN--N 121  
Db 86 AANKERAVLETAQOKNGKNMNLTTDKVAYQAGEAVHLNLTNLNTTSLAONTATVEVYS 145  
Qy 122 TENDNSQLKLTNNEQPSAATOANLKLNPQAKAVQNAK--IDAGSLTDDQINELNKNIF 179  
Db 146 LENLKLTLQYT-----KYLPPNESYTTQKGEFVIPANSLANNR-GYLLKVN 191  
Qy 180 SKS-----AEKGAKL-----TFKDLRGIG-----NAIVKQDPQYAIPIYNAKEIKN 220  
Db 192 SDSQNNILEQGNRAIATEDDWRFTFPYAAIGSGKDNNSVLTKNLPDY---YRELEQMKN 248  
Qy 221 MPAT-----YTV-----DA-----OTGKMAHLDV 239  
Db 249 MNINSYFFDYKTSATNPPNPKFDQSNWNSHQVETDAVKALVNRVHQTGAVAL-- 306  
Qy 240 WDSWPVQD-----PVTGYVSNTM--GYQLVIAMGIPNPTGDNH IYLLYN-----K 284  
Db 307 YNMILAQANETAVLPDTEYIYNYETGGY-----GQNGQVMTYSIDDKPLQ 352  
Qy 285 YGDNDIFS-HWRN--AGSTFGTKNVTNFWGSGSAIVNDGTLQLFPTSNDTSD-----YKL 337  
Db 353 YYNPLSKSWQNYISNAMAQAKTKGFGDQWQDGTI--GDNRV-LSHNQKDSRDIAHSFML 409  
Qy 338 ND-----ORLATATLNLNVDNGVSIKSDVN-----YQVL--PEGDGFHYQTYE 379  
Db 410 SDVYAEFLNKKMKELPQYYLTJLN-DVNGENISKLSKODVIYNELWPFGT SALGNRPOE 468  
Qy 380 OFANGKDRENDDYCLRDPHVQLENGDRYLV-----FEAN-----TGTEYD 420

Search completed: September 26, 2002, 18:26:47  
Job time: 186 sec

Db 469 SYGDLKARVDQ-----VRQATGKSLIVGAYMBEPKFDNRIPLNGAARDVLASATY 519  
Qy 421 QSDDQYNNWANYGGDDAFNPKSSFKLLNNKKDRELAGLAN-----GALGILKLTNNOSKPK 476  
Db 520 QTDVAVLLTTAAIAAAGGYHM-----SLAALANPNDGGGVGL----- 556  
Qy 477 VERVYSPVSTLMACDEVXXKLGDKYLFVSVTRVSRGSDRELTAkontivGDNV----- 530  
Db 557 -ETAYIPTQSL-----KVSKELNKNVHY-----QQFITAYEN-LLRDKVENDSAE 600  
Qy 531 -----AMIGVVSLSLMGKYKPLNNSGVVLTASVPANMRTATYSYAYVPVAGHPDQVL--- 582  
Db 601 PQTFtangrQLSODALG-----INGDQVWTYAKKGNDFRTI-----QLLNLM 642  
Qy 583 -TTSYMSNKDFASGEGNYATWAPSLVQINPDOTTTVLAR-----A 622  
Db 643 GITSWKNEED---GYENKNT-----PDEQTNLLVTVPLTGVSMAEADRIAKOVYL 689  
Qy 623 TNOGDVWDDSSRNDNMGLGVLEGAANSALPGEMGKPV-----DWSLINRSPG 671  
Db 690 TSPDDWL-----QSSMISLTQVKTN-----ENGDPVLYIQVPRUTLMDMIYINET-- 735  
Qy 672 LGLKPHQVPQPKIDQPDQOPSGQNTKNVTPGNGDKPAGKATPONTNIDPSAQPSGQNTNI 731  
Db 736 --IKPETPKVP--EQPOH-----PARTLEPAIPQTPPEAVSLPVANKQAEQNGK 780  
Qy 732 DPSAQX-SGQNTKNVTPGNEKQKNTDAKOLPOTGNKSGLAGLYAGSLLALFLGAAI 787  
Db 781 NELVSALTGBENDLQLPTLSKRSLSISQAELPOTGDNNET-----RSNLLKVIGAGAL 833





Db 403 PHVIEDNGDRYLVEASTCTENYQGEDQIYNFTNYGSSAYNVKSLERFLDODDMYNRA 462  
QY 457 GLANGALGILKLTNNQSKPKVEEYVSLVTLMACDEV-----XXKLGDKYLYFSVTRVSR 512  
Db 463 SWANAAGILKLGDKKTPEVDQFYPLLSSTVSDLEPNNVYKLGDKYLYFTASRLNH 522  
QY 513 GSDRELTAQDNTIVGDNVAMIGVYVSDSLMGKYKPLNNSGVVLVLTASVPANWRTATYSYAV 572  
Db 523 GSNDAWNAKANEVGDVNNVLMGLYVSDLTNGYKPLNNSGVVLVLTASVPADWRTATYSYAV 582  
QY 573 PVAGHPDVLITSYMSKNKDFASCEGYNATWAPSLVQINPDPTTTLARATNQGDDWDD 632  
Db 583 PVAGSSDTLLMTAYMTRNEVAGKGNSTWAPSLVQVLPDGTTKVLAENTQCGDWLWDE 642  
QY 633 SSRNDNMLGVLEKGAANSALPEWCKPVDWLSLNSPGLKPHQVPQKIDPQPOQPS 692  
Db 643 PSRTTDTVGTL-----DTAYLPENGDYIDNVIG---GYGLKPHPTPGQ-----YQPT 687  
QY 693 QQWT-----KNVTPGNGDKPAGKATPDNTNIDPSAQPSGQ 727  
Db 688 VPSTPIHTDDIIISFEVSFDGHLVLPKPVKNVNDSSAGRIDQSRNSGGSLNVAFNVA---GG 744  
QY 728 NTNIDPSAAXSGONTNVPNGEKOQKTDAKOLPQTGNKSGLAGLYAGSLIALF 782  
Db 745 NISVKFS-QKSHNTKET-----KKAHVSVTEKKQKGN-----FFAALLALF 787  
RESULT 2  
JQ0802  
leivansucrase (EC 2.4.1.10) precursor - Bacillus amyloliquefaciens  
N:Alternate names: 2,6-beta-D-fructan-6-beta-D-fructosyltransferase; sucrose  
C:Species: Bacillus amyloliquefaciens  
C:Date: 12-Mar-1993 #sequence\_revision 12-Mar-1993 #text\_change 15-Oct-1999  
C:Accession: JQ0802; S11739  
R:Tang, L.B.; Lenstra, R.; Borchert, T.V.; Nagatajan, V.  
Gene 96, 89-93, 1990  
A:Title: Isolation and characterization of leivansucrase-encoding gene from Bacillus amyl  
A:Reference number: JQ0802; MUID:91092506  
A:Accession: JQ0802  
A:Molecule type: DNA  
A:Residues: 1-472 <ST>  
A:Cross-references: EMBL:X52988; NID:g939333; PIDN:CAA3179.1; PID:g939334  
A:Experimental source: ATCC 23844  
C:Genetics:  
A:Gene: sacB  
C:Keywords: glycosyltransferase; hexosyltransferase  
F:1-29/domain: signal sequence #status predicted <SIG>  
F:30-472/product: leivansucrase #status predicted <MAT>  
Query Match 18.1%; Score 751.5; DB 2; Length 472;  
Best Local Similarity 38.0%; Pred. No. 2.2e-34;  
Matches 190; Conservative 88; Mismatches 171; Indels 51; Gaps 19;  
QY 144 NLKLNQAKAVQNAKIDAGSLTDQINELNKNFSAEKGAKLTFKDLLEGIGNAIVK 203  
Db 2 NEKKIVQATVLTFTALLAGGATQAFKENQKAY-KETGVSVTHRMLQIPKQ--Q 58  
QY 204 QDPQVAPYNAKEIKNPNATYVDAQTKMAHLDVMDSWPQDPVTGVSYNMGYQLVI 263  
Db 59 QNEKYQVPFQDSTIKN-----IESAKG---LDVWDSWPLQN-ADGTVAEYNGYHVVF 107  
QY 264 AMMGIPNSPTGDHNYLLYKNYKQNDPSHWRNAGSIFGKTETNPF-----QWVS 312  
Db 108 ALAGSPKD-ADDTSYMFYQKVGNSIDSWKNAGRVF--KDSKDFDNDPILKQDTQWS 164  
QY 313 GSAIVDDGTIQLFTSNDTSDYKLNDRLATATLNVDNDGVSYSKVDNVQVLFEDG 372  
Db 165 GSATFTSGKIRLFT--DYSGKHGKQSLTAQVNVSKSDDTUKINGVEDHKTIFDGDG 222  
QY 373 FHYQTYEAFAN-GKDRENDYCYLRDPHVVLQENGDRYLIVFEANTGTED-YQSDQIYNWA 430  
Db 223 KTYQNVQOQFIDEGNYTSGDNHTRDPHYVE-DKGHKYLVFEANTGTENGYGESLFNKA 281

QY 431 NYGDDAFNIKSSPKLLNNKKDRELAGLALGILKLTNNQSKPKVEEYVSLVSTLMA 490-  
Db 282 YYGCGTTFRRKESQKLOOSAKKRD-AELANGALGIIELNNDYT---LKKVKKPLTITSNTV 337  
QY 491 CDEV-----XXKLGDKYLYFSVTRVSRGSDRELTAQDNTIVGDNVAMIGVYVSDSLMGKYK 546  
Db 338 TDETERANVFMRNGKWLFT---DSRGSKMTI-----DGINSNDIYMLGYVNSLTGPKPK 390  
QY 547 LNSGVVLVLTASVPANWRTATYSYAVPVAGHPDVLITSYMSKNKDFASGEGNTATWAPSF 606  
Db 391 LNKTLGLVQMLGDPNDVFTTSHFAVPOA-KGNNVITSYMTNRGFF--EDKKATEGPSF 447  
QY 607 LVQINPDPTTTLARATNQG 626  
Db 448 LMNIKGNKTSVVKNSILEQG 467  
RESULT 3  
A25040  
leivansucrase (EC 2.4.1.10) sacB precursor - Bacillus subtilis  
N:Alternate names: sucrose 6-fructosyl-transferase  
C:Species: Bacillus subtilis  
C:Date: 16-Aug-1988 #sequence\_revision 20-Feb-1995 #text\_change 20-Jun-2000  
C:Accession: S07309; A25040; I39967; H69702  
R:Steinmetz, M.; le Coq, D.; Aymerich, S.; Gonzy-Treboul, G.; Gay, P.  
Mol. Gen. Genet. 200, 220-228, 1985  
A:Title: The DNA sequence of the gene for the secreted Bacillus subtilis enzyme levan  
A:Reference number: S07309; MUID:85295507  
A:Accession: S07309  
A:Molecule type: DNA  
A:Residues: 1-473 <ST>  
A:Cross-references: EMBL:X02730; NID:g40118; PIDN:CRA26513.1; PID:g732568  
A:Experimental source: Marburg  
R:Shimotsu, H.; Henner, D.J.  
J. Bacteriol. 168, 380-388, 1986  
A:Title: Modulation of Bacillus subtilis leivansucrase gene expression by sucrose and  
A:Reference number: A25040; MUID:87008406  
A:Molecule type: DNA  
A:Residues: 1-68 <SHI>  
A:Cross-references: GB:M4202; NID:g143485; PIDN:AAA22725.1; PID:g143486  
R:Fouet, A.; Arnaud, M.; Klier, A.; Rapoport, G.  
Biochem. Biophys. Res. Commun. 119, 795-800, 1984  
A:Title: Characterization of the precursor form of the exocellular leivansucrase from B  
A:Reference number: I39967; MUID:84178454  
A:Accession: I39967  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-11,1',13-62 <RES>  
A:Cross-references: GB:K01987; NID:g143483; PIDN:AAA22724.1; PID:g143484  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber  
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal  
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Hoisappel, S.; Hosono, S.; Hullo, M  
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Losono, S.; Lardino  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau  
Y., M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete  
Riegler, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadale, Y.; Sato, T.; Scanl  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se  
akeuchi, M.; Tamakoshi, C.; Tanaka, T.; Terpestra, P.; Tognoni, A.; Tosato, V.; Uchiya  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida  
A:Authors: Yoshikawa, H.F.; Zumstede, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili  
A:Reference number: A69580; MUID:98044033  
A:Accession: H69702  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-473 <KUN>  
A:Cross-references: GB:Z99121; GB:AL009126; NID:g2635827; PIDN:CAB15450.1; PID:g26359  
A:Experimental source: strain 168

## C:Genetics:

A:Gene: sacB  
C:Keywords: extracellular protein; glycosyltransferase; hexosyltransferase  
F:1-29/Domain: signal sequence #status predicted <SIG>  
F:30-473/Product: levansucrase #status predicted <MAT>

Query Match 18.0%; Score 746.5; DB 2; Length 473;  
Best Local Similarity 36.3%; Pred. No. 4.2e-34;  
Matches 185; Conservative 92; Mismatches 162; Indels 71; Gaps 18;

QY 144 NLKLNQAAKAVQNAKIDAGSITDDQINELNFKNSAERKAKLTFFDKLEGIGNAIVK 203  
DB 2 NIKKFAKQATVLTFTTALLAGGATQAFKTN-----QPKYETIGISH-ITR 48  
QY 204 QD-----POYAIPYNAKEIKHNPATYTVDAQTGMAHLVDWSPVQDPVTGYV 253  
DB 49 HMLQIPEQOKNEKYQVPEFDSSTIKNISSA-----KGLVDWSPVQDN-ADGTV 97  
QY 254 SNYMGVQLVIAWMGIPNSTGDNHIIYLLYKNGDNDPFSHRNAGSIFGKTENVF----- 308  
DB 98 ANYHGYHIVFALAGDPKN-ADDTSIYMFYQKVGETSIDSWKNAGRVF--KDSKFDANDS 154  
QY 309 -----QEMSGSAIYNDGCTIOLFTSNDTSYKLNQRLATATLNLVDDNGVSIKSDV 362  
DB 155 ILKQDTQEMSGSATFTSDGKIRLFYF--DFSGKHYGKQITLTAQVNVASDSSLNNGVE 212  
QY 363 NYQVLFEGDGFHYQYEQFAN-GKRENDYDCLRDHPHVQLENGDRYLVEFANTGTED-Y 420  
DB 213 DYKSLFDGDKTYQYQFIDEGNYSNGDNHILRDPHYVE-DKGHYLVEFANTGTEDGY 271  
QY 421 QSDQIYNYWANGGDDAFNIKSFLLNKKRELALGALGILKLTNNQSKPKVEEV 480  
DB 272 QGEESLFNKAYYKSTSPFQESQKLLQSDKKR-TAELANGALGMIELNDDYT---LKKV 327  
QY 481 YSPLVSTLMADEV---XXKLGDKYLFSTVRSRGSDELTAQNTIVGDNVAMIGYV 536  
DB 328 MPLASNTVTEIERANVFKNRWYLFDSRGSKMTIDGITSND-----IYMLGIVY 380  
QY 537 SDLSMGKYKPLNNSGYVLTASVPANWRTATYSYAVPVAGHPDQVLITSYMSKNDFASGE 596  
DB 381 SNSLTGPYKPLNGLVLMKMDLPNDVTTYSHFAVPA-QGNVVTITSYMTNRGFYADK 439  
QY 597 GNYATWAPSLVQINPDPTTVTLARATNQ 626  
DB 440 --QSTFAPSLNKGKTSVVKDSITLQ 467

## RESULT 4

H97118  
levansucrase [Imported] - Clostridium acetobutylicum  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C:Accession: F97118  
R: Noll, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A: Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum ATCC824  
A: Reference number: A96900; MUID:21359325; PMID:21359325  
A: Accession: F97118  
A: Status: preliminary  
A: Molecule type: DNA  
A: Residues: 1-489 <KUR>  
A: Cross-references: GB:AE001437; PIDN:AAK79737.1; PID:g15024742; GSPDB:GN00168  
A: Experimental source: Clostridium acetobutylicum ATCC824  
C: Genetics:  
A: Gene: CAC1772

Query Match 16.5%; Score 685.5; DB 2; Length 489;  
Best Local Similarity 33.5%; Pred. No. 1.1e-30;  
Matches 170; Conservative 101; Mismatches 167; Indels 69; Gaps 18;

QY 161 IDAGSLTDDQINELNFKNSAERKAKLTFFDKLEGIGN-----AIVKQDPOYAIP 211  
DB 16 ICSGIIITQOTFASTNDNN-----YKTYGVSHITRYNMKSIPBQNDLKEKVP 63  
QY 212 YSNAKEIKHNPATYTVDAQTGMAHLVDWSPVQDPVTGYVSNYMGVQLVIAWMGIPNS 271  
DB 64 QFNASTLKNATASAKGYD-KGNLIDLDVWDSWPVQNG-DGTIVANYHGYHIVFALAGDPKN 121  
QY 272 PTGDNHIIYLLYKNGDNDPFSHRNAGSIFGKTENV-----FQEMSGSAIYNDGDT 322  
DB 122 -QDDTSIYMFYQKVGETSIDSWKNAGRVF--KDSKFDANDS 180  
QY 323 IQLFFT-----SNDTSYKLNQRLATATLNLVDDNGVSIKSDVQVLEF-GD 371  
DB 181 VRLFYDTSGVAKDGTAS-----NQVITTVQNLNQLSDPNTINIDSVSHKSVFDEGN 235  
QY 372 GFHYQYEQFAN-GKRENDYDCLRDHPHVQLENGDRYLVEFANTGTED-YQSDQIYNYW 429  
DB 236 GTIYQYQFIDEGNYSNGDNHILRDPHYVE-DKGHYLVEFANTGTEDGYQSDTSLNK 294  
QY 430 ANYGDDAFNIKSFLLNKKRELALGALGILKLTNNQSKPKVEEVYSLVSTLM 489  
DB 295 AFYGRSOSFFKTEKQDQLLIDTNKKHDASLALGALGILNNDYT---LKKEMKPLIASNT 351  
QY 490 ACDEV---XXKLGDKYLFSTVRSRGSDELTAQNTIVGDNVAMIGYVSDSLMGKYK 545  
DB 352 VTDIRANVFKNRWYLFDSRGSKMTIDGITSND-----IYMLGFSNLTGPYK 404  
QY 546 PLNNSGYVLTASVPANWRTATYSYAVPVAGHPDQVLITSYMSKNDFASGEGNTATWAPS 605  
DB 405 PLAKTGLVNLNLPDPTTVTLARATNQ 632  
QY 606 FLVQINPDPTTVTLARATNQGDWVDD 632  
DB 462 FLLNKGKTSVVKDSITLQ 488

## RESULT 5

H97118  
levansucrase [Imported] - Clostridium acetobutylicum  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C:Accession: H97118  
R: Noll, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A: Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum ATCC824  
A: Reference number: A96900; MUID:21359325; PMID:21359325  
A: Accession: H97118  
A: Status: preliminary  
A: Molecule type: DNA  
A: Residues: 1-428 <KUR>  
A: Cross-references: GB:AE001437; PIDN:AAK79739.1; PID:g15024744; GSPDB:GN00168  
A: Experimental source: Clostridium acetobutylicum ATCC824  
C: Genetics:  
A: Gene: CAC1774

Query Match 7.6%; Score 315; DB 2; Length 428;  
Best Local Similarity 27.1%; Pred. No. 3.1e-10;  
Matches 105; Conservative 68; Mismatches 158; Indels 56; Gaps 18;

QY 236 HLDVWDSHPVQDPVTGYVSNYMGVQLVIAWMGIPNSPTGDNHIIYLLYKNGDNDPFSHRN 295  
DB 66 NLWVMDTWPVKK-DGSLAVVNGYKVIIFALTASRVGNKRDVAGISYFCSTGDENWY 124  
QY 296 AGSIFGKTENVFQEMSGSAIYNDGDTIOLFTSNDTSYKLN--DORLATATLNLVDD 353  
DB 125 KGLAVNVEDALGSRQWAGSAILDENGVMQFFYATGRKEAVRTFEQRLVTKFSINVDK 184  
QY 354 NGVTSKSDYNDVQVLEFGDGFHYQYEQFANGKRENDYDCLRDHPHVQLENGDR-YLVE 412  
DB 185 GGVHTNCKSHQVILEPDGVYVQTMQQ-AKGP-----IYVFRDPYFFEDPKTKKDYLF 239

```

QY 528 DNYAMIGYVSDS-LMGKKYPLNNSGVLTASVPANRRTATSYAVPVAGHPDVLITY 586
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 307 ---GLYFVSENGIFGYEPLNGSLVLGG--PASPQTEAYAHYIM-----NNGLVESF 355
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 587 M-----SNKDFASGEGNYATWAPSFVLQVINDPT 615
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 356 INEIIDPKSGKVLAG-----SLAPTVPVLEQGHET 386
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT      7
S47527
extracellular sucrose - Zymomonas mobilis
C:Species: Zymomonas mobilis
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 17-Mar-1999
C:Accession: S47527
R:Song, K.B.; Lee, S.K.; Joo, H.K.; Rhee, S.K.
Biochim. Biophys. Acta 1219, 163-166, 1994
A:Title: Nucleotide and derived amino acid sequences of an extracellular sucrose gene
A:Reference number: S47527; MUID:94368848
A:Accession: S47527
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-413 <SON>
A:Cross-references: EMBL:L08094

```

Query Match 7.08; Score 292; DB 2; Length 413;  
Best Local Similarity 26.3%; Pred. No. 5.6e-09;  
Matches 104; Conservative 69; Mismatches 153; Indels 70; Gaps 20;

[illegible]

```

RESULT      8
JC2519
levansucrase (EC 2.4.1.10) precursor - Zymomonas mobilis
N;Alternate names: sucrose 6-fructosyltransferase
C;Species: Zymomonas mobilis
C;Date: 13-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 20-Jun-2000
C;Accession: JC2519; PC2376
R;Kyono, K.; Yanase, H.; Tonomura, K.; Kawasaki, H.; Sakai, T.
Biosci. Biotechnol. Biochem. 59, 289-293, 1995
A;Title: Cloning and characterization of Zymomonas mobilis genes encoding extracellular
A;Reference number: JC2519; MUID:95218269
A;Accession: JC2519
A;Molecule type: DNA
A;Residues: 1-423 <KYO>

```

QY 453 RELAGLANGILKLTNNOSKPKVEEVYSPVSTLMACDEVXXK---LGDKYFLFSVT 508

Query Match 5.8%; Score 242.5; DB 2; Length 423;  
Best Local Similarity 24.1%; Pred. NO. 3.2e-06;  
Matches 115; Conservative 76; Mismatches 173; Indels 113; Gaps 25;

QY 509 RVSRGSDRELTAQDNTVGDVAMIGYVSDSLMGKKYKPLNNSGVVLTASVPANWRTATYS 568  
Db 305 -----HKYTFADNLGTGDGV--YGFVSDKLTGTPTMNSSGLVL--GNPSSQPFQYYS 353  
QY 569 YTAVPVAGHPDQVLITSYMSN-----KDFASGEGNYATWAPSLFVQINPD 613  
Db 354 HYVMP-----NGLVTSFSDSPMKGKDYRIG----GTEAPTVKILLKGD 393

RESULT 11  
B97033  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C:Accession: B97033  
R:Noiling, J.; Breton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Bacteriol. 183, 4823-4836, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: B97033  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2817 <KUR>  
A:Cross-references: GB:AE001437; PIDN:AAK79053.1; PID:gl5023993; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC1079

Query Match 4.9%; Score 203.5; DB 2; Length 2817;  
Best Local Similarity 21.7%; Pred. No. 0.0078;  
Matches 143; Conservative 92; Mismatches 211; Indels 213; Gaps 35;

QY 34 NNYNGVAEVT-----ERQA--NGOIGVDGKIISANTSTSGTNOESSA--TNNT 80  
Db 66 NNLTKVQTNATSSMPNTNPKATNNSKILVNPKNQASSNEGTPPKQASIPYTNVT 125  
QY 81 ENAVNNSKNTNTENAVNE-----NKNTNNT-----ENAVNENKNTNTEN 124  
Db 126 DN-----KNTFKNESSINNEAPIPKDTSKTSQAOTKGSNDNNIPSNNTSTYSKN 179  
QY 125 D-----NSOLKLT-----NNEQPSAATOANLKLNPQAKAVQNAKIDAGSLTDD 169  
Db 180 ENPSNTDIKTEAPANAPKDTPNNSDLSALAKNALSNNLNLAADSSQTSKVTSSN--- 235  
QY 170 QINELANKINFSKAERKALTFKDLLEGIGNAIVKQDQPAIYPSNAKEIK-----NMPAT 224  
Db 236 --NDAPKVNNTSTDKASNLNDSQDG---WTKDGKKY--YVNGVQKGFQSIKNSIY 288  
QY 225 YTVD---AOTGKAHLVDWDSHPVQDPVGYVSNYGYQIVIAMMGIPIPSPTGDHNIYLL 281  
Db 289 YFNDDGSMQTG-----W-----LKYNSSNYFDASGVML-----TGLQNINGT 326  
QY 282 YNKYGDNDFSHRNAGSIFGTKETNVFQBSGSAIVNDGTLQ--FFTSNDPDSYKLNQ 340  
Db 327 Y--YGEND-----DGKLLTGLQAIN-----NNYFENNDGVMTGWTTCNDSKIYFDNNG 374  
QY 341 RLATATLNLNVDDNG-----VSIKSDNYQVLFEG-----DGPHY----- 375  
Db 375 VMOTGLVHNNKYGYFGNDGKLLTGLQINNTYTYFDSNGVMQTDWITIDGSKYFVSNG 434  
QY 376 --QT-----YEOFANGKRENDYCLRDHPVQLENGDRYIVFEAN----- 414  
Db 435 VMOTGIIISYGYGFA-----NGKLLTG---LQVINGNSY-YFDTNGIRLVSRTITI 484  
QY 415 TGTEDYQSDQIY--NWANYGGDDAFNKSFKLLNNKKDLRELACALGAILKLTNQ 472  
Db 485 DGKDYFNDQGITLWNWYNDKYYFYISG-----VQOTGLQ 521  
QY 473 SKPKVEEYVPLVSTLMACDEYXXLGDYKYLFSVTRVSRGSDRELTAQDNTVGDN-VA 531

Db 522 N-----IDGNYFFDSSGIMQTLQKIDGK--TYFFGDNGIR 556  
QY 532 MIGYVSDSLMGKKYKPLNNSGVVLTASVPANWRTATYSYAVPVAGHPDQVLITSYMSNK 590  
Db 557 QLGWITYO--NKKYY--FNSDGSQMOTDLKIYSYSTSPYNYHYQYGFDFNDGKLLTGLQTIK 613

RESULT 12  
AH2515  
hypothetical protein alr7304 [imported] - Anabaena sp. (strain PCC 7120) plasmid pCC  
C:Species: Anabaena sp.  
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002  
C:Accession: AH2515  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kurlitz, T.; Sasamoto, S.; Watanabe, A.; Irigi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. (strain PCC 7120)  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AH2515  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-4936 <KUR>  
A:Cross-references: GB:BA000020; PIDN:BAW78388.1; PID:gl7135842; GSPDB:GN00180  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: alr7304  
A:Genome: plasmid

Query Match 4.9%; Score 203.5; DB 2; Length 4936;  
Best Local Similarity 20.4%; Pred. No. 0.018;  
Matches 185; Conservative 123; Mismatches 320; Indels 277; Gaps 45;

QY 6 KMWAVATLSASILMGVYTAHADQVESNNY-----NGVAEVNTERQANGQIGVDGKIISA 61  
Db 413 QNTAALTIVNDLPPQGTINFNI-----NNYTVNENGTSINILVRTG---GSDGEV--S 461  
QY 62 NSNTTSGSTNOESSATNTENAVNENKNTNTENAVNENKNTNTENAVNENKNTNTEN 121  
Db 462 VTLTPSDGDTAGSDYNNLPITVTFANGTSKTIN--LISQNGLFFDGDYNDVNDPANFSE 520  
QY 122 TENDNSQLKLTNNEQPSAATOANLKLNPQAKAVQ--NAK-----I 161  
Db 521 TK-DTFTIELWAN---PTATRAS---TPETSGVNAFFENQKIYFPPKQGLTGLTSNDV 572  
QY 162 DAG-----SLTDDQIINELANKINFSKAERKALTFKDLLEGIGNAIVKQDQPAIYPS 213  
Db 573 YAGISGTGVTISBHTLNYMPSVLVYNTA-----LSGWNHIALVYENKTPKLYI 622  
QY 214 NAKEIK--NMPATYTVDAQT-----GKMAHLVDWDSWPVQDPV----- 249  
Db 623 NQOFTKAGLTSQIYVHPSLSFGGTSIROQDWSFKGSDIDVRIWHKARTKEEIKAGLNREL 682  
QY 250 TGYVSNYMGYQLVIAAMGIPNSPTGDHNIYLLYKYGDN---DFSHRNAGSIFGTKETN 306  
Db 683 TGNESGLIG-----YWNFSINGINIVODLSTNNKNGTFFGQAQSTA 722  
QY 307 VFOEWSGSAIVND-----GTQLFTSNDPDSYKLNQDLATATLNLNVDDNGYSIKSV 361  
Db 723 GF---STSFINDNIYEPIETVNLTLT--NPTGGATLGTOK---TANLIVNDNDIACTI 774  
QY 362 D-----NYQVLFEGDGFHYQTEQFANGKRENDYCLRDHPVQLENGDRYIVFEANTGT 417  
Db 775 QFSNANYAVENGTAVNAVTLNR--TNGSD---GVVSVR-----INLTNG-----TATAG 819  
QY 418 EDYQSDDOIYWNANYGGDDAFN-----IKSSFKLNNKKDLRELACALGAL-----GIL 466  
Db 820 SDYNNSPITVNFAD--GENSKVTIPIIDDSILESINESINLTLANPTNGATIGTQNSAVV 877  
QY 467 KLTNNOSKPKV-----BEVYSPIVSTLMACDE-----VYXX 496  
Db 878 NIIDNKLPLTLVMTABQLTEGNTIOCTVTRNTDTTTEPLTVTLVNSDNTQITVPTVTI 937

```

Qy 497 KLGDKYILFSVTRVSRGSDRELTAKDNTI-VGDNVAMI-----GYVSDSLMGKYKPLNNSG 551
Db 938 PAGANVNFSTAVD-----DNLIELPRNYSIIASAPGFISGS--DSVGVIDNDA 985
Qy 552 VYLTAASVPAANWRTATYSYAVPVAGHPDQVLITVSMN--KDFASGEGNYATWAPSLVQ 609
Db 986 VTLSLTVDVT-----TWINMGGRAIATITRNIVTDIP-LVVQ 1021
Qy 610 INPDQTT--TVLARATNOGDWYDSSRNDNMLGLVLRKEGAANSAALPGEWGKPVWDWSLIN 667
Db 1032 LSTSDTTEATVPATVI-----AAQASATFEI-QGVDDTIVD 1058
Qy 668 RSPGLGLKPHQPQVPKIDOPQPSGONTKNVTPGNGDKPAGKATPDNTNIDPSAQPSG- 726
Db 1059 GTQAV-IITARPIYTNV--AVPTGNATANLVVDNESPSPSKLTIDRLISETGTATAI 1115
Qy 727 --QWNTIDPS-----AQXSGQNTKNVTPGNEKQKGNKTDARLPQTG 765
Db 1116 ITRNTNTDSALVVTLNSSDTTEATVPNTVTIAAGOTSATFTITGVSDGINDSSQNVITA 1175
Qy 766 NKSGL 770
Db 1176 AANGL 1180

RESULT 13
T28680
fibrinogen-binding protein homolog - Staphylococcus aureus
C:Species: Staphylococcus aureus
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-May-2000
C:Accession: T28680
R:Josefsson, E.; McCrea, K.; Ni Eidhin, D.; O'Connell, D.; Cox, J.; Hook, M.; Foster, T.
Microbiology 144, 3387-3395, 1998
A:Title: Three new members of the serine-aspartate repeat protein multigene family of St
A:A:Reference number: Z20510; MUID:99098700
A:Accession: T28680
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1166 <JOS>
A:Cross-references: EMBL:AJ005647; NID:el318793; PID:el318794; PIDN:CAA06652.1
C:Genetics:
A:Gene: sdre

```

Query Match	4.88;	Score 199.5;	DB 2;	Length 1166;
Best Local Similarity	22.0%;	Pred. NO. 0.0035;		
Matches 166;	Conservative	99;	Mismatches 312;	Indels 177; Gaps 39;

```

12 TLVSASTLMGGV-----TAHADQ-----VESNNYNGVAEVT 44
   : : : : : : : : : : : : : : : : : : : : : : : :
29 TVGTASILVGTTLIFLGLNGQEAKAENTSTENAKQDDATTSDNKEVVSEYENNSTYNN 88
   : : : : : : : : : : : : : : : : : : : : : : : :
45 ERQANGIGVDGKLIISANSTTSGSTNQESSATNTENAVNVNKSNTNTENAVNVNKN 104
   : : : : : : : : : : : : : : : : : : : : : : : :
89 TNPKEKNTNDSOPEAKKESTSSSQOQNVTATTETKPNIEK-----ENVKPSDTKT 143
   : : : : : : : : : : : : : : : : : : : : : : : :
105 TTNTENAVNVNENKNTNTENDNSQLKTNNQEPSAATQANLKLNPQAAKAVON-----158
   : : : : : : : : : : : : : : : : : : : : : : : :
144 ATEDTSVILEEKAAPNTNNDVTTKESTS--EPS-TSEIQTKPTTPESTNIENSOPQT 200
   : : : : : : : : : : : : : : : : : : : : : : : :
159 -AKIDAGSLTDOQINELINKINFSKSAEKGAKLTFKDL-EGIGNAIVKQDPOVAIPYSNAK 216
   : : : : : : : : : : : : : : : : : : : : : : : :
201 PSKVD--NQVTDATNPKEPVNYSKEELKNPEKLKELVRNDSNTDHSITKPVATAPTSVAP 258
   : : : : : : : : : : : : : : : : : : : : : : : :
217 EIKNMPATTVYDQTKGMAHLVD-----SWPVQDPVPTGVSYNMGYQLVIAMGP 269
   : : : : : : : : : : : : : : : : : : : : : : : :
259 KRYNAKRRFAV-AQPAAVASNNYNDLIKVTQTKIKVGDGKDNVAAAHDKGDI-----EYD 312
   : : : : : : : : : : : : : : : : : : : : : : : :
270 NSPTGDHNHLYLVNKKYGDN-DFSHWRNA--GSIFGCTKETNVQEWSSGAIVN---DDGTI 323
   : : : : : : : : : : : : : : : : : : : : : : : :
313 TEFTINKV-----KKGDTWTINDKNVIPSLDTDKNDIDITDSGSEVIKAGTDEKAKP 367
   : : : : : : : : : : : : : : : : : : : : : : : :

```

```

324 QY QLFFTSNTSD-----YKLNQDQ-----LATA-----TLNLNVDDNGVS 357
368 Db QITTYTDYDKYEDIKSRLTLYSIDKKTVPNETSILNTFATAGKETSQNTYDYQDPM 427
358 QY IKSVDNQTQVLEGGDFHYQYEQ--FAN--CKDREN-----DDYCLRDPHVVVQLE 403
428 Db VHGDSNTQSITFKLDEKDQTEQQIYVNPULKSATNTKVDIAGSQDDY-----GNIKLG 482
404 QY NGDRYLVEAFENTGEDYQ--SDDQ-----IYNWANGGDADFNIKSFKLLNNKK--DR 453
483 Db NGS--TIIDQNTKEIKYKVNSDQQLPQSNRIYDFSQY-----DVTISQF--DNKRSFSN 532
454 QY ELAGLANGALGILKLTNNQSKPKVEEYSPLV--STLMACDEVXXKIGDKYILFSVTRVSR 512
533 Db NVATLDFGDI-----NSAYIIKVVSKYTPDSGELDIAQGTSMRTTDKYGYNYA--- 582
513 QY GSORELTAKONTIVGDNA-----MIG-YV-----SDSLAG---KYPLNNSGWLTA 556
583 Db GYSNFIVTSDGGDGTVPPEKLYKIGYVWEDVDKGVOGTQSKERPMAN--VLVTL 640
557 QY SVP-----ANRWTATYSYVAVPAGHDQVLI-----TYSMSNKDFASGSGNTATWAPSFL 607
641 Db TYPDGTTKSVRTDANGHYEFGGLKDGETTVVKFETPTGLPKVNGCTTDGKDSNGSSVT 700
608 QY VOINPD-----DFTTVLARATNOGDVWDDSSRN 636
701 Db VKINGKDDMSLDTGFYKEPKYINGLYVWEDTNKD 734

```

RESULT  
S22737

glucosyltransferase (EC 2.4.1.1.) S - Streptococcus salivarius  
C:Species: Streptococcus salivarius  
C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 21-Jan-2000  
C:Accession: S22737; S28810; B44811; S22727  
R:Jacques, N.  
submitted to the EMBL Data Library, March 1992  
A:Reference number: S22726  
A:Accession: S22737  
A:Molecule type: DNA  
A:Residues: 1-1599 <JAC>  
A:Cross-references: EMBL:Z11872; NID:g47530; PIDN:CAA77898.1; PID:g47531  
A:Experimental source: ATCC 25975  
R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.  
J. Gen. Microbiol. 137, 2577-2593, 1991  
A:Title: Molecular characterization of a cluster of at least two glucosyltr  
A:Reference number: A44811; MUID:92148377  
A:Accession: S28810  
A:Molecule type: DNA  
A:Residues: 1-51 <GIF>  
A:Cross-references: EMBL:Z11873  
C:Genetics:  
A:Gene: gtfK  
C:Superfamily: cpl repeat homology  
C:Keywords: glucosyltransferase; hexosyltransferase  
F:1456-1475/Domain: cpl repeat homology <CPR>

Query Match	4.8%;	Score 198.5;	DB 2;	Length 1599;
Best Local Similarity	19.6%;	Pred. NO. 0.0064;		
Matches 179;	Conservative 109;	Mismatches 304;	Indels 323;	Gaps 45;

Qy	1	MYKVGKNW---	AVATLSASILMGVVTAH-	ADQVESNNTNG-----	VAEVNTERQANG	50
		: : : :	: : : :	: : : :	: : : :	
Db	9	LHKVKQWYTLAVASVALATIVGSGVATSSLASAEETNNNSGSPSTTVGE-NINPVVEK	67			
Qy	51	QIGVDGKILIAANSNTT-----	SGTNGESSATNTNNAVNESKNTNTNTNNAVV	99		
		: : : :	: : : :	: : : :	: : : :	
Db	68	EVGTTTAVANTSNAITTTTERRAEVTDADKPGATTVQPSNGCTTSDRAAAVEVEAKPETTAKPEV	127			
Qy	100	NENKNTNTNNAVYNNKNTNTNTNDNSQLKLTNNQEPSAATQANLKK-----	147			
		: : : :	: : : :	: : : :	: : : :	
Db	128	ATRPETATTSEVANAGVAAPTEKSK- --ELSEAEETKAAVSLDNTKCKDGKYYLLED	184			



Query Match 4.7%; Score 195.5; DB 2; Length 1093;  
Best Local Similarity 19.5%; Pred No. 0.0054;  
Matches 169; Conservative 122; Mismatches 236; Indels 341; Gaps 50;

QY 148 -----LNQAAKAVONAKIDAG-----SLTDQINELNKNINFSKSAEAGKLTFFKDL 194  
DB 185 GSHKKNFAITVNGQVLYPDENGALSSSTSYSTFTQETTLNLTDFTKNNAAYDSTRASFELV 244  
QY 195 EGINAIVKQDPOYAIYPSNAKEIKNPNATYTVDAQTGKMAHLVDWSPVQDPVTVGVVS 254  
DB 245 DGYLTA-----DSWY-----RPKELEAGTWKATEKDFRPLLSW--WPKOTQVAYL- 292  
QY 255 NYMGYQL-----VIAMGIPNSPTGDN-----HIYLLYNYKGD----- 287  
DB 293 NYMTKALNGEBETKDVETIENSQASLAAQAQILQKIEVKIAANKSTDWLRQSTAEAFVKD 352  
QY 288 -----NDFS-----HWRNAGSIFCTKTNVFOEWSGS-----AIYN-DD 320  
DB 353 QDKWNIINSESPGKHFQKALLFVNSDST--KWANSDYRKLNQATSYIKNHKIVNGSD 409  
QY 321 GTIQLTSTNDTSD-----YKLNQDLATATLNLNVDDNGVYSIKSVDNVQV 366  
DB 410 GGYE-FLLSNDIDNSNPVQAEMLNQLYIFPMWQIIVFGDKDKDAHFDGIRVDAVDNVSV 468  
QY 367 LFEGDGFHYQYFQFANGKRENDYCYL-----RDPHVY-----OLENGDR 407  
DB 469 -----DMLQVSSYKAAKYVNESEARALANISILEAMSHNDPYVNEHNTAALSDMNGLR 524  
QY 408 YLVFE-----ANTGT-----EDYQSDDOIYNWANY----- 432  
DB 525 LSIYVHGLTRPTVNGTGNASMKDLINGYFGLSNRAEVTSDQL-GFYATYLFVRAHDS 583  
QY 433 -----GGDAF-----NIKSSFLLNKKDKRELALANGALGILKLTNNQ 472  
DB 584 EVQTVIADIISKIDPTDGTFTLQDKQAQFIYN-----ADMLKVDKEY 629  
QY 473 SKPKVEEYVPLSTLMACDEVXXKLGDKY-----YLFSTVRSR----- 512  
DB 630 THSNIPAYALMLOTWGAATRVY--YGDLYTDNGQYMAKSPYFDQITLLKARPKYVAG 687  
QY 513 -----GSDRELAKONTIVGDNVIMGYVSDSLM-----GRYKPLNNSGVV-LTA 556  
DB 688 GQTSYIHNLAGDGVSSAKDNKEY---LVSVRYGQDLSKTDTEGGKYG--RNSGMLTLIA 742  
QY 557 SVP-----ANWRATYSYAYVPVAGHPDQVL-----ITSYMNKD-----FASGEG 597  
DB 743 NNPDKLADGETIVNM-----GAHRNQAYRPLLLGTEKGIYSSLSNDSPTKIYKYTDAQG 798  
QY 598 N-----YATWAPSFVLQINPDDTTVLA-----RATNOGDWYWDSS 634  
DB 799 NLVFTADEIKGFKTVMDSGYLSVWVP-----VGATDQNVLAKPSTKAYKEGDKVYSSA 853  
QY 635 RDNMLGVLYKEGAAN 649  
DB 854 ALEAQ--VIYEGFSN 866

RESULT 15  
B86748  
hypothetical protein ykbc [imported] - Lactococcus lactis subsp. lactis (strain IL1403)  
C:Species: Lactococcus lactis subsp. lactis  
C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
C:Accession: B86748  
R:Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, Q.; Malarne, K.; Weissenbach, J.; Ehrlich  
Genome Res. 11, 731-753, 2001  
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s  
A:Reference number: A86625; MUID:21235186; PMID:11337471  
A:Accession: B86748  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-1093 <STQ>  
A:Cross-references: GB:AE005176; PID:g12723930; PIDN:AAK05084.1; GSPDB:GN00146  
A:Experimental source: strain IL1403  
C:Genetics:  
A:Gene: ykbc

Search completed: September 26, 2002, 18:25:13  
Job time: 92 sec





Db 27 QNSPQKVPQFNASAIKIDSAGYD-KSGNLIDLVDWSDWPLQN-ADGTAANYHGYHIV 84  
QY 263 IAMGIPNPTGDHIIYLLNKNYKNDPFSHRNAGSIFGTRETWV-----FOEWSG 313  
Db 85 SALAGDPKN-SDDTPLHLFYQKVGDTSIDSMKNAGRFEEDMKFVPNDPILKYQTQEWG 143  
QY 314 SAIVNDGTTIQLFTSNDTSDYKLN-----DQRLATATLNLVDDNG-VSIKSDVN 363  
Db 144 SATLTKGQVRLFT-----DYSNPDGGTGAGNQIISTAQVNLSDPDAATLKVGVSD 198  
QY 364 QYVLE-FDGPHGYQTYEQFAN-GRDRENDYCLDRPHVQVLENDRIYLFNANTGTE-Y 420  
Db 199 HKSVEFDGDTVIYNIQFIDEGKWSIDNHLRDPHVE-DKGHKYLVFNANTGTTDGY 257  
QY 421 QSDQIYNWYAGGDAFNKSSFKLNKKDLRELALANGALGILKLTNNQSKPKVEEV 480  
Db 258 QGDSFNKAYYGGSDVFFQNEKNKLLQSPK-KQIASLANGALGIVELADDT---VKSV 313  
QY 481 YSPLVSTLMACDEV---XXKLGDKYVLFVTRVSRGSDRELTAKNITVGDNVAMIGYV 536  
Db 314 MKPLVASTVADEVERANIFKMNKWLFTDSRGSKMTSGINDKD-----VYMLGPG 366  
QY 537 SDSLMGKYKPLNNSGVVLTASVPANWRTATSYAYVPVAGHPD--QVLITSYMSKNDF 592  
Db 367 GDSLNGPHNPINETGLVLNNLDPADLTHYSHCGIP---HPEGNNVYLTSYMTNRF 421

RESULT 2  
US-09-503-172A-2  
; Sequence 2, Application US/09503172A  
; Patent No. 6284510  
; GENERAL INFORMATION:  
; APPLICANT: ITO, Tetsuya  
; APPLICANT: FUJITA, Koki  
; APPLICANT: HARA, Kozo  
; APPLICANT: TONOSUKA, Takashi  
; APPLICANT: SAKANO, Yoshiyuki  
; TITLE OF INVENTION: BETA-FRUCTOFURANOSIDASE GENE  
; FILE REFERENCE: 10749-0001-0  
; CURRENT APPLICATION NUMBER: US/09/503,172A  
; PRIOR FILING DATE: 2000-02-14  
; PRIOR APPLICATION NUMBER: JP 160416/1999  
; PRIOR FILING DATE: 1999-06-08  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 578  
; TYPE: PRT  
; ORGANISM: Arthrobacter sp.  
US-09-503-172A-2

Query Match 6.4%; Score 267; DB 4; Length 578;  
Best Local Similarity 23.8%; Pred. No. 1.2e-12;  
Matches 124; Conservative 82; Mismatches 196; Indels 118; Gaps 26;  
QY 200 AIYKQDPQYAIYPSNAKEIKNMPATYV-----DAOTGKMAHLDVWDSPVQDPVTGVVS 254  
Db 67 ADARQLQMSDPTAPSR-NSMPASVTMTPTVPQDFPDMSNEQVQVWDTWPTLDEANQYS 125  
QY 255 NYNGYQLVIAWGIPIPSPTGDNHYY-----LLYKNYK-----DNDFSHRNAGSIFGTK 303  
Db 126 -VNGWEIIFSLVADRNLGDFDRHIVFAKIGFYRPAAGVPAARPEN--GGWTYGLGVPEK 182  
QY 304 ET-NVFO-----EWSGSAIVNDGTTIQLFT-----SNDTSDYKLDORLATATLN 348  
Db 183 VTQIPEQDSFSHTQWSGARYSKNGEIKLFTDFVAFYNSOGTNIKPDPRIALSVGK 242  
QY 349 LNVDDNGSVKSDVNYQVLFEGDFHYQTYEQFANGKRENDYCLDRPHVQ--LENGD 406  
Db 243 VKANKKGVLTGFKVKTDLQADGTYQTGAQ-----NEFFNFRDPTFEEDPAHPGE 294  
QY 407 RYLVEFANTGTEDYQSDQIYNWYAG---GD-----DAFNKSKFKLNNKKREL 457

Db 295 TFMVFEGNSA---MQRETATCNEADLYRQGDPAEYAEVDDVNASGATYQIGN----- 343  
QY 458 LANGALGILKLTNNQSKPKVEEVYSPVSTLMACD-----EVXKXKLGDKYVLFVSYTR--- 509  
Db 344 -----VGLAKAKNKOLTE--WEFLPILISANCVTQOTERPOIYFKDG-KSYLFTISHRGT 395  
QY 510 VSRGSDRELTAKNITVGDNVAMIGYVSDSLMGKYKPLN-NSGVVLTASVPANW----- 562  
Db 396 FAAGLD-----GPEGVYGVGDGIRSDYQPLNGSGSLALGNPTNLNFGGQPF 443  
QY 563 -----RTATSYAYVPVAGHPDQVLLITSYMSKNKDFASGEGNYATWAPSEFLVQI--N 611  
Db 444 APDFNQHPGHFQAYSHYMP--GGLVQSFIDTIGTHDDFVRG-----GTLAPTVMKDIGVG 497  
QY 612 PDDTTTTLARA--TNOGDVWVDDSSRDNMGLVYKKEGAANS 650  
Db 498 GDPYKTAVDYSYSGSLGSGWADIPANKHLFTNGKFGVAVS 537

RESULT 3  
US-08-381-936-2  
; Sequence 2, Application US/08381936  
; Patent No. 5792923  
; GENERAL INFORMATION:  
; APPLICANT: ROBER, Manuela  
; APPLICANT: GEIER, Gebhardt  
; APPLICANT: GEIDER, Klaus  
; APPLICANT: WILLMITZER, Lothar  
; TITLE OF INVENTION: DNA sequences which lead to the  
; TITLE OF INVENTION: formation of polyfructans (levans), plasmids containing  
; TITLE OF INVENTION: these sequences as well as a process for preparing  
; TITLE OF INVENTION: transgenic plants.  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen  
; STREET: 1180 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: US  
; ZIP: 10036-8403  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/381,936  
; FILING DATE: 09-FEB-1995  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 93 02110  
; FILING DATE: 09-AUG-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE P42270618  
; FILING DATE: 08-DEC-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meilman, Edward A.  
; REGISTRATION NUMBER: 24,735  
; REFERENCE/DOCKET NUMBER: P/951-108  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 382-0700  
; TELEFAX: (212) 382-0888  
; TELEX: 216925  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 415 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-381-936-2

[illegible]

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Baron, Ronald J.
; REGISTRATION NUMBER: 29,281
; REFERENCE/DOCKET NUMBER: 294-29
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 822-3550
; TELEFAX: (516) 822-3582
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-08-362-232-2

Query Match 4.7%; Score 197; DB 1; Length 543;
Best Local Similarity 23.8%; Pred. No. 3.3e-07;
Matches 117; Conservative 78; Mismatches 175; Indels 122; Gaps 31;

QY 220 NMPATYTVDAQTGKMAHLD--VWDSWPVQDPVTGYVSNMGYQLVIAMGIPNSPTG--D 275
D 78 NIPADFPV-----INPDVWVWDTLIDKHADQFS--YNGWEVIFCLTADPNAGYGFDD 129
QY 276 NHLY-----LLYK-----YGDNDFSHRNAGSIFGKTETNVFQWESGSAI 316
D 130 RHVHARIGFFYRRAGIPASRRPVNGWTYGGHLFPDGASAQVYAGQTYTNO--AEWSGSSR 188
QY 317 VND--DGTIQLFFT-----SNDTSYKLN--DQRLATATL--NLNVDDNGYSIKSVQVLF 368
D 189 LMQIHGNTVSVEYTDVAFNRDANANNITPPQAIITQTLGRIHADFNHVVFTGFTAHTPL 248
QY 369 EGDGFHYQTYEQFANGKRENDNDYCLDRPHVQ--LENGDRYLVEANT-----GTEDIYS 422
D 249 QPDGVLYQ-----NG--AQNEFFNFRDPTFEDPKHPGVNMYVFEAGTORGVANCTE 300
QY 423 DDQIYNWANYGGDDAFNIKSFLLNNKKDRELA--GLANGALGILKLTNNQSKPKVEEV 480
D 301 AD-----LGRPNDP--NAETLQEVLDGAYYOKANIGLA-----IATDSLKWK---F 345
QY 481 YSPLVSTLMACDEVYXKL-----GDKYILFSV---TRVSRGSDRELTAKNITVGDNVAMI 533
D 346 LSPILISANCVNDQTERPQVYLHNGKYYIFTISHTTTAAAGVD-----GPD-GVY 393
QY 534 GYVSDSLMGYKPLNNSGVLTASVPANWRTA-----TYSYAVPVAG 576
D 394 GFVGDGIRSDQPM--NYGSLTGNPTDLNTAAGTDFDPSDQNPRAFQSYSHYVMPGG- 451
QY 577 HPDQVLITSYMSNKKDFASGEGNYATWAPSLVQINPDQTTTTLARATNOGDWVDD--SS 634
D 452 -----LVESFIDTVENRRG-----GTLAPTVRRI--AQNASAVDLRYNGGLGCGYDIPAN 501
QY 635 RND-NMLGVLKE 645
D 502 RADVNIAGFIQD 513

RESULT 6
; Sequence 2, Application US/08814196
; Patent No. 5731173
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Fructosyltransferase Enzyme, Method
; TITLE OF INVENTION: For its Production and DNA Encoding the Enzyme.
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann & Baron
; STREET: 350 Jericho Turnpike
; CITY: Jericho
; STATE: New York

```

```

; COUNTRY: United States of America
; ZIP: 11758
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk, 3.5 inch, 1.44 Mb
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 6.0 for DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/814.196
; FILING DATE: 10-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362.232
; FILING DATE: 22-December-1994
; APPLICATION NUMBER: CU 125/93
; FILING DATE: 23-December-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Baron, Ronald J.
; REGISTRATION NUMBER: 29,281
; REFERENCE/DOCKET NUMBER: 294-29
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 822-3550
; TELEFAX: (516) 822-3582
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-08-814-196-2

```

```

Query Match 4.7%; Score 197; DB 1; Length 543;
Best Local Similarity 23.8%; Pred. No. 3.3e-07;
Matches 117; Conservative 78; Mismatches 175; Indels 122; Gaps 31;

```

```

QY 220 NMPATYTVDAQTGKMAHLD--VWDSWPVQDPVTGYVSNMGYQLVIAMGIPNSPTG--D 275
D 78 NIPADFPV-----INPDVWVWDTLIDKHADQFS--YNGWEVIFCLTADPNAGYGFDD 129
QY 276 NHLY-----LLYK-----YGDNDFSHRNAGSIFGKTETNVFQWESGSAI 316
D 130 RHVHARIGFFYRRAGIPASRRPVNGWTYGGHLFPDGASAQVYAGQTYTNO--AEWSGSSR 188
QY 317 VND--DGTIQLFFT-----SNDTSYKLN--DQRLATATL--NLNVDDNGYSIKSVQVLF 368
D 189 LMQIHGNTVSVEYTDVAFNRDANANNITPPQAIITQTLGRIHADFNHVVFTGFTAHTPL 248
QY 369 EGDGFHYQTYEQFANGKRENDNDYCLDRPHVQ--LENGDRYLVEANT-----GTEDIYS 422
D 249 QPDGVLYQ-----NG--AQNEFFNFRDPTFEDPKHPGVNMYVFEAGTORGVANCTE 300
QY 423 DDQIYNWANYGGDDAFNIKSFLLNNKKDRELA--GLANGALGILKLTNNQSKPKVEEV 480
D 301 AD-----LGRPNDP--NAETLQEVLDGAYYOKANIGLA-----IATDSLKWK---F 345
QY 481 YSPLVSTLMACDEVYXKL-----GDKYILFSV---TRVSRGSDRELTAKNITVGDNVAMI 533
D 346 LSPILISANCVNDQTERPQVYLHNGKYYIFTISHTTTAAAGVD-----GPD-GVY 393
QY 534 GYVSDSLMGYKPLNNSGVLTASVPANWRTA-----TYSYAVPVAG 576
D 394 GFVGDGIRSDQPM--NYGSLTGNPTDLNTAAGTDFDPSDQNPRAFQSYSHYVMPGG- 451
QY 577 HPDQVLITSYMSNKKDFASGEGNYATWAPSLVQINPDQTTTTLARATNOGDWVDD--SS 634
D 452 -----LVESFIDTVENRRG-----GTLAPTVRRI--AQNASAVDLRYNGGLGCGYDIPAN 501
QY 635 RND-NMLGVLKE 645
D 502 RADVNIAGFIQD 513

```

QY 203 KODPQAIYPSNAKEIKNNPARYTYVDAQTGKKAHLVDWDSNPVQDPVTGYYSNVNGYQLV 266  
 Db 383 LPSGDYIL-----KEIE-APRPYTFDKK-----EYPTMKDFDNOGY--- 419  
 QY 263 IAMMGIPNSPTGDNHILLYNKYGDNDFSHWRNAGSIFGKTETNVFQEWSGSAIYVNDGDT 322  
 Db 420 -----FTTIENAKAIEKTKDVSQAQWEGTQKV----- 447  
 QY 323 IQLFTSNDTSYKLNQRLATATLNLNVDDNGVSKSVNDVNQVILFEGDGHFYQYEOFA 382  
 Db 448 -----KPTIYFKLYKQ-----DDN-QNTTPVDKAEIKKLEDGTTTKVTSNLP 488  
 QY 383 NGKDRENDYCLDRPHVVQLENGD--RYLVFEANTGCTEDYQSDDDQIYNWANYGGDDAFNI 440  
 Db 489 -----END-----KNGKAIKYILVKEVNAQGEDTTPGYTKK----- 519  
 QY 441 KSSFKLNNKKRELALANGALGILKLTNNOSKPKVEEIVSPL-----VSTLMACDEV 494  
 Db 520 ENGLVNTNEKPIETIYSISGEKVWDDK--DNQDGKRPEKVSYNLLANGEKVKTL-----DV 573  
 QY 495 XXKLGDKYILFSVTRSVRSRSDRELTAKNNTIVGDNVAMIGYVSD-----SLMGKYKPLNNS 550  
 Db 574 TSETNKKYEFKOLPKPYDEGKKIEV-----TVTEDHVK--DYTTDINGTTITNKYTPGETS 626  
 QY 551 GVVLTASVPANWRATYISYAYVPVAGHPQOVLITSYMSNKDFASGE-----GNYA-TW 602  
 Db 627 -----ATVTKNWDNNNQ-----DGRKRPEIKVELYQDGK--ATGKTALINESNNHTHW 674  
 QY 603 -----APSELVQINPDQTTTVLARATN-----QGDWYVWDDSS 634  
 Db 675 TGLDEKAGQOVKYTVBELTKVAGYTHVDNNDMGNLIVTNKYTPETTSISGEKVWDD-- 732  
 QY 635 RNDNMLGVLKEGANSAAALPGEMGKPYD-----WSLINSRPGILGPKHPQVPQPKIDQP- 687  
 Db 733 -KQNDQGRPEKVSYNLLADGEKVKTLDTVSETNW-----KYEFKOLPKPYDEGK 780  
 QY 688 --DOQPSQNTKNTVPG-NGDKPAGKATPONTNIDPSAQSGQNTNIDPSAQSGQNTKN 744  
 Db 781 KIEYTVTEDHVKYTTDINGTTITNKYTPGET-----SATVTKN 819  
 QY 745 VTPNEKOGKN-TDAK-OLPOTGNKSG 769  
 Db 820 WDDNNQDGKRPEIKVELYQDGKATG 846  
 RESULT 8  
 US-08-293-728-2  
 ; Sequence 2, Application US/08293728D  
 ; Patent No. 6008341  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Foster, Timothy J.  
 ; APPLICANT: McDevitt, Damien L.  
 ; TITLE OF INVENTION: The S. aureus Fibrinogen Binding Protein Gene  
 ; FILE REFERENCE: 05344.105011  
 ; CURRENT APPLICATION NUMBER: US/08/293,728D  
 ; CURRENT FILING DATE: 1994-08-22  
 ; NUMBER OF SEQ ID NOS: 20  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 2  
 ; LENGTH: 933  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus aureus  
 ; US-08-293-728-2

US-09-421-868-2

QY 66 --TSGSTNOESSATNTNPNENAV-----VNESKNTNNT--ENAVVNE--NKNTNTNTENA 111  
Db 69 DDINVSDTKSSNTNNGETSSAQNPAQOQETQSSSTNATTEETPVTGEATTTTNOANTP 128  
QY 112 VVNEKNKNTNTENDNSOLKLTNNQPSAAQANLKLNPQAAKAVQNAKIDAGSLTDQOI 171  
Db 129 ATTQSSNTNABEELVQ-----TSNETTFENDTNTVSSVSPONSTNAENV-----STTQDTS 179  
QY 172 NELNKNFSAEKGAKLTPEKLEGIGNATVKODPOQVAIPYSNA-----KE 217  
Db 180 TEATPSN--NESAPQSTDAENK-----VNVQAVNTSAPRMAFSLAAVAADAPAGATD 231  
QY 218 IKNPATYTVDAQTKMAHLVDWDSWPQDPVTGYVSNYMGYQLVIAMMGIPNSPT--GDN 276  
Db 232 ITNQLTNTVTGIDSGTIV-----YPHQ---AGYVKLVNGF-----SVPNSAVKGD 274  
QY 277 HIYLLYKNGDNDFSHRNAGSIFGTETNTVFOEWSGSAIVNDDGTTLQFTSNDTSDYK 336  
Db 275 F-----KITVPKELNL-----NGVTSTAK 293  
QY 337 L-----NDORLATATLNLNVDNGVSIKSDNYQVLFEGDGFHYQYEQFANGKDRDND 391  
Db 294 VPPIMAGDOVLA-----NGV-----IDSDGNVIYTFDYVNTKD--DVK 330  
QY 392 YCLRDPHVQLENGDRYLVEANTGTEDYQSDQIYNWANYGGDDAFNIKSFLLNNKK 451  
Db 331 ATLTPAYI-----DPENVKKT----- 347  
QY 452 DRELALANGALGILKLTNNOSKPKVEEYSPVSTLMACDEVXXKLGDKYLFVSVTRVS 511  
Db 348 -----GNVTLATGISTTANKTVLDYEKG-----KFYNLSI----- 380  
QY 512 RGSDELTAKTNTI-----VGNVAMIGVSDSLMGKYKPLNNSGVLTASVPANMR 563  
Db 381 KGTIDQIDKNTNTYRQTIYVNPBGDNV-----IAPVLGNLKPNTDSNALI-----DQ 429  
QY 564 TATYSYYAVPVAGHPDQVLTISYMSN--KDFASGEGNYATWAPS---FLVQIN--PDDTTTV 618  
Db 430 NTSIKVYKVDNAAD-----LSESYFVNPFENFEDVTNSVNTTFPNQYKVEFNTPDQIT 485  
QY 736 QXSGQNTKN 744  
Db 580 DSGSDSTSD 588

RESULT 9  
US-09-421-868-2  
; Sequence 2, Application US/09421868  
; Patent No. 6177084  
; GENERAL INFORMATION:  
; APPLICANT: Foster, Timothy J.  
; APPLICANT: McDevitt, Damien L.  
; TITLE OF INVENTION: The S. aureus Fibrinogen Binding Protein Gene  
; FILE REFERENCE: 05344.105011  
; CURRENT APPLICATION NUMBER: US/09/421,868  
; CURRENT FILING DATE: 1999-10-19  
; PRIOR APPLICATION NUMBER: 08/293,728  
; PRIOR FILING DATE: 1994-08-22  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus

Query Match 4.2%; Score 173; DB 4; Length 933;  
Best Local Similarity 20.5%; Pred. No. 6.le-05;  
Matches 162; Conservative 90; Mismatches 273; Indels 264; Gaps 40;

QY 6 KNAVATLVASILMGVTVTAHADOVESNNYNGVAEVTNERQANGQIGVGGKIISANSNT 65  
Db 14 KSGIVASVLVGTLLGFLGSLKESKADASENSVTSQSDASNESKSN-----DSSSVSAAPT 68  
QY 66 --TSGSTNOESSATNTNPNENAV-----VNESKNTNNT--ENAVVNE--NKNTNTNTENA 111  
Db 69 DDINVSDTKSSNTNNGETSSAQNPAQOQETQSSSTNATTEETPVTGEATTTTNOANTP 128  
QY 112 VVNEKNKNTNTENDNSOLKLTNNQPSAAQANLKLNPQAAKAVQNAKIDAGSLTDQOI 171  
Db 129 ATTQSSNTNABEELVQ-----TSNETTFENDTNTVSSVSPONSTNAENV-----STTQDTS 179  
QY 172 NELNKNFSAEKGAKLTPEKLEGIGNATVKODPOQVAIPYSNA-----KE 217  
Db 180 TEATPSN--NESAPQSTDAENK-----VNVQAVNTSAPRMAFSLAAVAADAPAGATD 231  
QY 218 IKNPATYTVDAQTKMAHLVDWDSWPQDPVTGYVSNYMGYQLVIAMMGIPNSPT--GDN 276  
Db 232 ITNQLTNTVTGIDSGTIV-----YPHQ---AGYVKLVNGF-----SVPNSAVKGD 274  
QY 277 HIYLLYKNGDNDFSHRNAGSIFGTETNTVFOEWSGSAIVNDDGTTLQFTSNDTSDYK 336  
Db 275 F-----KITVPKELNL-----NGVTSTAK 293  
QY 337 L-----NDORLATATLNLNVDNGVSIKSDNYQVLFEGDGFHYQYEQFANGKDRDND 391  
Db 294 VPPIMAGDOVLA-----NGV-----IDSDGNVIYTFDYVNTKD--DVK 330  
QY 392 YCLRDPHVQLENGDRYLVEANTGTEDYQSDQIYNWANYGGDDAFNIKSFLLNNKK 451  
Db 331 ATLTPAYI-----DPENVKKT----- 347  
QY 452 DRELALANGALGILKLTNNOSKPKVEEYSPVSTLMACDEVXXKLGDKYLFVSVTRVS 511  
Db 348 -----GNVTLATGISTTANKTVLDYEKG-----KFYNLSI----- 380  
QY 512 RGSDELTAKTNTI-----VGNVAMIGVSDSLMGKYKPLNNSGVLTASVPANMR 563  
Db 381 KGTIDQIDKNTNTYRQTIYVNPBGDNV-----IAPVLGNLKPNTDSNALI-----DQ 429  
QY 564 TATYSYYAVPVAGHPDQVLTISYMSN--KDFASGEGNYATWAPS---FLVQIN--PDDTTTV 618  
Db 430 NTSIKVYKVDNAAD-----LSESYFVNPFENFEDVTNSVNTTFPNQYKVEFNTPDQIT 485  
QY 619 LARATNQGDWVDSSRNDNMLGVLKEGNASAAALPGEWGKPVWDS---LNRSPGLGLK 675  
Db 486 PYIVVVNGH--IDPNSKGLDALRSTLYG--YNSNII---W-RMSWDNEVAFNNGSGSDG 538  
QY 676 PHQPVPKIDQDQOPSGQNTKNVTGNGDKPAGKATPDNTNIDPSAOPSGQNTNIDPSA 735  
Db 539 IDKPVPV--EQPDE-----PGEIE-----PIPEDSDSDPGSD--SGSDSNSDSSG 579  
QY 736 OXSGQNTKN 744  
Db 580 DSGSDSTSD 588

RESULT 10  
US-09-268-347-30  
; Sequence 30, Application US/09268347  
; Patent No. 6335182  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS  
; FILE REFERENCE: 1038-860  
; CURRENT APPLICATION NUMBER: US/09/268,347

;; CURRENT FILING DATE: 1999-03-16  
;; NUMBER OF SEQ ID NOS: 54  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 30  
;; LENGTH: 1004  
;; TYPE: PRT  
;; ORGANISM: Haemophilus Influenzae  
US-09-268-347-30

Query Match 4.2%; Score 173; DB 4; Length 1004;  
Best Local Similarity 20.6%; Pred. No. 6.9e-05;  
Matches 191; Conservative 122; Mismatches 333; Indels 280; Gaps 46;

Qy 5 GKNMAVATL-----VSASILMGVVTAAHADOVE-SNNYNGV--AEVNTERQANGQ 51  
Db 162 GHHTVTFLEKDLNVKNATVSDKLSLG---ANGNKVDITSDTNGLKFAKPTSTNGQ-NGN 216  
Qy 52 IGVDGKIISANSNTTSGSTNQESSATNTE-----NAVVNESK 89  
Db 217 VHLNG-IASLTDTITGTT---KSATNGVDVQNHNRASVADVLNAGNINQNGASVDFV 272  
Qy 90 NTNNTENAVVNENKNTNENAVVNENKNTNTENDNSOLKLTNNEQPSAAQANLKL 149  
Db 273 NYTDVDFVNGLNTNVNTTFAHNR-KTTRVVDVTGLPVQVTEDEGTIVKVGNEYEA 331  
Qy 150 POAKAVONAKIDAGSLTDDQI-----NELNKNFSKSA---EKAKLTFKDELGIGNAI 201  
Db 332 KODGSADMDKKVENGKLAATKVKVLSANGTNPVKSINADGTENTDAVSFKQKAL---- 387  
Qy 202 VKDPOYAIPIYNAEIKMNPATYTVDAQTGRMAHLVDWSPVQVPTGYVSNYMGYQL 261  
Db 388 --QDKQVTLASNAY-----ANGSSDADGGK-----GIQT-----LSNGLNFKF 424  
Qy 262 VIAMMGIPSPGTHNHYLYNKYGNDPFSHRNAGSIRGTRTNVQFWSGSAIVNDG 321  
Db 425 -----KSTGGE-----LNIKAEND-----TWTFTPKKGSVQVGDG 456  
Qy 322 --TIQ-----LFTSHTSDYKLNQORLATATNLNVD-----NGVSIKS 360  
Db 457 KATIQDGAATTTGLVBESLNLKGLKVGTDGTGTGTDHTDILVSGDKVTLKA 516  
Qy 361 VDNQVLPFGDGFHYQTYEQPANGKRDRENDYC-----LRDPHVQLENGDRYL 410  
Db 517 GNLKVKQEGTFTYALKDELTDVKSVEPKDTANGANGASTIKDKGLTTPANGAG--A 574  
Qy 411 FEANTG-----TED-----YQSDDOLYNNWYGGDDAFN-IKSSFLLNKKDRELGLA 459  
Db 575 AGANTANTISVTKDGISAGNKAVNVVSLKKFGDANFDPDLPSSADNLTQYDNYKGLT 634  
Qy 460 N-----GALGILK-----LTNNSKPKVEEVSPLVSTLMACDEVXX 496  
Db 635 NLDEKSKGQTPTVADNTAATVGLRGLGWVISADTKGELNKEYNAQVRN---ANEVXF 691  
Qy 497 KLGDKYLLSVTRVSGRSDRELT---AKDNTIVGDNVAMIGYVSDSLMGKYKPLNNSGVV 553  
Db 692 KSGNGINVSQKT-LDNGT-REITFEALD-----ENAIAPGSGSKALR-----DNTVAI 738  
Qy 554 LTASVPANRTATYSYVAVPVAGHPDVLITSYMSHK---DFASGEGNTATWAPSELVQI 610  
Db 739 GTGNVNAEKSAGF-----GDP-----NVIDKAGSGYAFGDNDRIT----- 775  
Qy 611 NPDDTTTVTLARATNOGDWVDDSSRNDMLG-----VLKEGAANSALPGEWKPDVMSL 665  
Db 776 --SKNTFVLGNSVNA-----KRDANGNVLTBEKEVVGKDGAKTKVTVPQALGETVENS 827  
Qy 666 -----INRPSGLGLKPHQVOPKIDQPPQPSGQNTKNVTTCGNDKPKAGKATPDNTNI 718  
Db 828 YLGNASTATKDKGNLK-----SDGTAGTNTTATAGATGTVNGFAGATAHGAHSV 875  
Qy 719 DPSAQP-----SGQNTNIDPSAQ--XSGQNTKNVTTCGNEKOGKNTD----- 757  
Db 876 GASGEERRIONVAAGEISATSDAINGSQLYAVAKGVNTNLAGQVNVKGRKADAGTASALA 935

Qy 758 AKQLPOTGNKSGLAGLYAGSLALFG 783  
Db 936 ASQLPQ-----ASMPGKSMVSIAG 954

RESULT 11  
US-08-446-855A-2  
; Sequence 2, Application US/08446855A  
; Patent No. 5849573  
; GENERAL INFORMATION:  
; APPLICANT: Stewart, Thomas S  
; APPLICANT: Flores, Maria V  
; APPLICANT: O'Sullivan, William J  
; TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl  
; TITLE OF INVENTION: phosphate synthetase II  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon & Vanderhye PC  
; STREET: 1100 No. 5849573th Glebe Road, 8th Floor  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22201-4714  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/446,855A  
; FILING DATE: 06-Jul-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mitchard, Leonard C  
; REGISTRATION NUMBER: 29,009  
; REFERENCE/DOCKET NUMBER: 47-80  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-816-4000  
; TELEFAX: 703-816-4100  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2391 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-446-855A-2

Query Match 4.2%; Score 173; DB 2; Length 2391;  
Best Local Similarity 20.3%; Pred. No. 0.00028;  
Matches 115; Conservative 93; Mismatches 227; Indels 132; Gaps 28;

Qy 31 VESNNYNGVAEVTERRQANGQIGVDGKIISANSNTTSGST-----NOESSATNTNTENAV 84  
Db 1275 VKNHLYNEVDV-DKDTQLH-----KENNNNNMNSGNVENCKLNKESIGYNNSSNCI 1326  
Qy 85 VNESKNTN--NTENAVVNE---NKN-----TNNTENAVV--NENKNTNNT----- 122  
Db 1327 -----NTNNINIENNICHDISINKNIKTINNNSNNSINNENVTNLCVYSERAGSHIY 1381  
Qy 123 -----ENDNSOLKLTNNEQPSAATQANLKLNPQAAKAVQNAKIDAGSLTDDQNEL 174  
Db 1382 GKBEKSGSDDTNLSAQSNNNFSCNNNNKAN-----VDVNVLEND-TKKR 1429  
Qy 175 NKINFSASGAKLTKFDELGIGNAIVKODPQVAPYPSNAKEIKNMPATYTVDAOTGKM 234  
Db 1430 EDINTTTTVFEGQNSVINNNKKNSSLLKGDDEEDIVVMVNLKKNENYNSVINNVDCRCKDM 1489  
Qy 235 AHLDVDWSPVQDPTGVYSN-YMGYQL---VIAMMGIPSPGTHNHYLYLYNKYGNDNF 290  
Db 1490 -----DGKINDECKTYKKNKYKMDGLNINNIVDELNGTSHSTNDHLYLDNFTSDEEI 1543

QY 291 SHWRNAGSIFGKTETNVFQWGSALVNDGGTIOLEFFTSNDTSYKLLNDORLATATLNLN 350  
Db 1544 GNNKNN-DWYLSKESISNKNPGNSVYVVDVY-----NNEYKINKMKELIDNENLN 1594  
QY 351 VDDNGVSKSVNYQVLFEGDGFHYQTYEQFANGKDR-----ENDDYCLRD-----PHVV 400  
Db 1595 -----DEYNNVNMNCSNYNNASAFYNGKDRNDNLEND--CIEKNMDHYTKHYN 1641  
QY 401 OLEN-----GDRYLVEANTGTEDYQSDQIYNWANYGGDDAFNKSFFKLLNNKKDRELA 456  
Db 1642 RLNNRSTNERMMLAVNNEKESNHEKGHR-RNGLNKNKE-----KNMEKNKGKNDKKNY 1696  
QY 457 GIANGALGILKILTNOSKPKVBEVYSPLVSTLMACDEVYXXKLGDYKYY-----LFSVTRVS 511  
Db 1697 HYVNH-----KRNEYNSNIEKFNYY-----DDINKK---EYEDENDIYYFTSS 1742  
QY 512 RGSDELTAKDNTIYVGNVAMIGYVSD 538  
Db 1743 QGNNDL-SNDNYLSSELTDEYDD 1768

## RESULT 12

US-09-150-741-2  
; Sequence 2, Application US/09150741  
; Patent No. 6183996  
; GENERAL INFORMATION:  
; APPLICANT: Stewart et al.  
; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate  
; Patent No. 6183996  
; TITLE OF INVENTION: Synthetase II  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/09/150,741  
; CURRENT FILING DATE: 1998-09-10  
; EARLIER APPLICATION NUMBER: PL6380  
; EARLIER FILING DATE: 1992-12-16  
; EARLIER APPLICATION NUMBER: AU93/00617  
; EARLIER FILING DATE: 1993-12-02  
; EARLIER APPLICATION NUMBER: 08/446,855  
; EARLIER FILING DATE: 1995-07-06  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO. 2  
; LENGTH: 2391  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: protein  
US-09-150-741-2

Query Match 4.2%; Score 173; DB 4; Length 2391;  
Best Local Similarity 20.3%; Pred. No. 0.00028;  
Matches 115; Conservative 93; Mismatches 227; Indels 132; Gaps 28;  
QY 31 VESNNGVAEVTNRQANGIGVDGKIIISANSNTTSGT-----NOESSATNNTENAV 84  
Db 1275 VKNHLNVEVD-DKDTQLH-----KENNNNNMNSGVNCKLNKESGYNNSSNCI 1326  
QY 85 VNESKNTN--NTENAVYNE---NKN-----TNNTENAVY--NENKNTNT----- 122  
Db 1327 -----NTNNINIENNICHISINKNIKVTNNSSNNSNENVTNLCVSEKSHIY 1381  
QY 123 -----ENDNSOLKLTNNEQPSAQTQAMKLNPOAAKAVQNAKIDAGSLTDQINEL 174  
Db 1382 GKEKSGSDDTNLSAQSNNNFSCNNMKNAN-----VDNVNLEND-TKKR 1429  
QY 175 NKINFSKSAKGAKLTFKDLEGIGNAIVKQDPOVAIPYSNAKEIKNMPATVTVDAQTKM 234  
Db 1430 EDINTTTVFMEGQNSVNNKKNKESNLSLLKGDEEDIVVNLNKKENNSVINNVDCRKDM 1489  
QY 235 AHLDWDSWPQDPVTGVSN-YMGYQL-----VIAMMGIPNSPTGDNHIYLLNKGNDNF 290

## RESULT 13

US-08-169-927-2  
; Sequence 2, Application US/08169927  
; Patent No. 5783441  
; GENERAL INFORMATION:  
; APPLICANT: Carl, Mitchell  
; APPLICANT: Dobson, Michael E.  
; APPLICANT: Ching, Wei Mei  
; APPLICANT: Dasch, Gregory A.  
; TITLE OF INVENTION: Gene and protein applicable to the  
; TITLE OF INVENTION: Preparation of vaccines for Rickettsia prowazekii and  
; NUMBER OF INVENTION: Rickettsia typhi and the Detection of Both  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Counsel, Naval Medical R & D Command  
; STREET: Bldg. 1, T-12, 8901 Wisconsin Ave.  
; CITY: Bethesda  
; STATE: MD  
; COUNTRY: USA  
; ZIP: 20889-5606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/169,927  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/742,128  
; FILING DATE: 08/09/91  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Spevack, A. David  
; REGISTRATION NUMBER: 24,743  
; REFERENCE/DOCKET NUMBER: 75,976  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 295-6759  
; TELEFAX: (301) 295-1022  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1612 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-169-927-2

Query Match 4.1%; Score 172; DB 1; Length 1612;  
Best Local Similarity 19.5%; Pred. No. 0.00018;



Matches 200; Conservative 117; Mismatches 366; Indels 342; Gaps 47;

```
QY 4 VCKNNAVATLVASILMGVVTAAHQDVESNNYNGVAEVNTER-----QANGQIG 53
Db 245 VSKNGA-ATEFNVTGLTG-----NLKIIELNTAAVAGKLISLGAANAVIG 291
QY 54 VD-----GKIIS--ANSNTTSG-----STNOESAT-----NNTENAV 84
Db 292 TDNGAGRAAGFTVSDNGNAATISQVYAKNVIQSANAGGQVTFEHIHVDVGLGTTNFK 351
QY 85 VNESKNTNTNENAVNENKNTNT-----ENAVNENKNTNTE-----123
Db 352 TADSK-----VIITENSGSTNFGNLDQIVVPDRTKILKGNFIDGVKNNGNTAGVIT 404
QY 124 -NDNSQLKTNNEOPSATQAN-----144
Db 405 FNANGALVASTDPNIATVNTNATEAGAGVVELSGIHAELRLNGGSGIFKLADGTVIN 464
QY 145 -----LKKLNPOAKAVQ--NAKIDAGSLTDQIN--ELNKINFSKGAERKAKLTFKDL 194
Db 465 GPVNONALMNNALAAAGSIQDLSAIIITGDIGNGVNAALQHIITLANDASKILADGANI 524
QY 195 EG--IGNAIVKQDPQYAIPIYSNAEIKNPATYTVDAQTGMALHDVWDSPVQD--PVTG 251
Db 525 IGANVGGAHFQANGGTIKLTNTQ--NNIVVNFDLIDITDKTVGVVDASLNTNQTTLING 582
QY 252 YVSNMVGQYLVAMGIPNSPT--GD-----NHILYLYNKYGDNDPSHWRNA 296
Db 583 SIGTVANTKTLAGNLGSSKTLINAGDVAINELVIENNGSVQLNHNHTYLTIKTINAANQ 642
QY 297 GSI-----EGTKETNY-----PQWSSGSAIV 317
Db 643 GQIIIVAADPLNTNTLADGTNLGSAENPLSTIHFATRAANADSLINVGKGVNLYANNITT 702
QY 318 NDDGTIOLFSTNDTS-----DYKLNDRQALATATLNLVDDNGVSTKSVDNTQVL 367
Db 703 NDANVGSILHFRSGGTSIVSGTVGQQGKHLNLL-----DNGTTVK-----744
QY 368 FEGDGFHYQYEQFANGKRENDYDCLRDHPVQLENG--DRYLVFEANTGTEDYQSDQ 425
Db 745 FLGD-----TTF-----NGGK-----TEGKSILQISNYYTTHVESADNTGTLEFVNDP 790
QY 426 I-----YWNANG-----GDDAFNTKSPKLLN-----NKKRELALANG-----461
Db 791 ITVTLNKQGAIFGVLYKQVLIISGPGNIVFENGVIVHGIANAANSISPENASLGTSFLPLS 850
QY 462 --ALGILKLTNNQSKPVEEYSPVSTLMACDVXXK--LGDKYLYLFSVTRYRSGSDR 516
Db 851 GTPLDVLTIKSTVNGTVDNFNAPIV--VVSIGDMSINNGQIIGDK--NIIALSIGSDN 906
QY 517 ELTAKDNTIYGDNVAMTIGYVSDSLMGKYKPLNNSG--VVLTAASVPAN-----561
Db 907 SITVNANTLYS-----GIRTTKNNGTVTLSGGHPNPGTIIYGLGLENGSP 952
QY 562 -WRTATY-----SYAYVPVAGHPDQVLITWSMKNDF-----ASGEGNYATWAP 604
Db 953 KLQVFTFTDYNLNGSIANNVTINDVTLTTGGIAGTFDAKITLGSVNGNAV-----1007
QY 605 SFLVQINPDPTTVLARATNOG-----WVWDDSSRDNDMLGLVLEKGAANSALPGE- 656
Db 1008 RFVDSTFSDPRSMIVATQANGVTYVIGNALVSNIGSLDTPFVASVRTPGNDSGAGLQGN 1067
QY 657 WGPVPDW-----SLINRSPGLGLKHPQVOPKIDQDPQPSGQNTKNVTPGNDKPKAGKA 711
Db 1068 YSQNIDFGTVNLILSNVILG--GGTTAINGEID-----LLTNLIFANGSTWGDN 1118
QY 712 TPNTNIDPDAQSGQ---NTNIDPDAQSGQNTKNVTPGNEKQGNKNTDAKQLPQTGNK- 767
Db 1119 TSISTTLNVSNGIGQVIAEDAQVNATTGTTTIKIQDNANANFSGTQAYTLIQGGARF 1178
QY 768 SGLAG 772
Db 1179 NGTLG 1183
```

```
RESULT 14
US-09-336-447A-9
; Sequence 9, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP41 AND USP42 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 941
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-9

Query Match 4.1%; Score 170.5; DB 4; Length 941;
Best Local Similarity 20.7%; Pred. No. 9.8e-05;
Matches 190; Conservative 108; Mismatches 380; Indels 239; Gaps 41;
```

```
QY 4 VCKNNAVATLVASILMGVVTAAHQDVESNNYNGVAEVNTERQANGQIGVQD-----KI 58
Db 111 ICKNSTV-----GGFTNEA---MGEYSTVAG--GANNQAKGNTSTVGGGNGKA 155
QY 59 ISANSNTTSGSTNQ-----ESSATNTNENAVNENKNTNTNENAVNENKNTN 106
Db 156 IGNNSTVVGSSNNQAKGHESTTAGCKNNOATGNGSFAAGVENKADAN--NAVALGNKNTI 213
QY 107 NTENAVNENKNT-----NNTENDNSQ---LKLTNNEQPSAATQANLKLN--149
Db 214 EGTNSVATGSSNTVTGKENVFILGNTNTENAQSGVLLGNNTAGKAATTVNNAEVNGL 273
QY 150 -----PQAAKAVON--AKIDAGSLTDQINELNKNFSK---SAEKAKLTFDLEGIGN 199
Db 274 TLENFAGASKANANNIGTVSGS-----ENNERQIVNVGAGQISATSTDAVNGSQLHALAK 329
QY 200 AIVKODPOYAIPIYSNAEIKNM--PATYTVDAQTGKMA-----HLVDWDSWPVQDPVTGY 252
Db 330 AVAK-----NKSDIKGLKGVKELDKREVGLSRDINSLHDDVDAN---QDSIAKN 376
QY 253 VSNYMGYOLVITAMM---GIPNSPTGDNHIIYLYNKYGDNDPSHWRNAGSIFG-----301
Db 377 KADTKGLNKEVKELDKREVGLSRDIGS-----LHDDVDADNODSIAKNADIKGLNKEVKE 431
QY 302 -TKETNVFOEWSGSAIVNDGDTIQLFFTSNDTSDYKLNDRQALATATLNLN---VDDNGVS 357
Db 432 LDKEVGLSRDIGS--LHDDVATNQADIAKNOADIKTLNENVEEELNLSGLRIDQKADI 489
QY 358 IKSVDNYOVLFEFGDGFHYQYEQFANG-----KRENDYDCLDRPHVVQ 401
Db 490 DNNINNIYELAQDQDQHSDDIKTLKNVVEEGLLSDLSGLRIDQKADLTDKITLKN---545
QY 402 LENG-----DRYLVFEANTGTEDYQSDQI---YWNANGYGGDDAFNIK 441
Db 546 VEEGLLDLSGLRIDQKADIAKNOADIAQNOTDIQDLAAYNELQDYAQKQTEAIDALNKA 605
QY 442 SS-----FKLLNNKDRDELALANGALGILKLTNNQSKPKVEEYSPVSTLMAC 491
Db 606 SSANTDRIATAEGLTAENKDKAQIA-----KAQANENKDI-----AK 643
QY 492 DEVXXKLGDY-----YLFVSTRYRSGSDRELTAKDNTIYGDNVAMIG-----534
Db 644 NQADIQLHDKKITNLGLIHSVAVRANGVNTGCVATNKAQADIAKNOADIANNTKNTIYELAAQ 703
```

QY 535 ---YVSD-SLMGKYKPLNNSGVVLT-ASVPANWRTATYSYAVPVAGHPDQVLITSYMSN 589  
Db 704 ODQSSDIKTAKVSAANTDRIAKNAEADASPETLTKNQNTLIEQG---EALVEQNKA 760  
QY 590 KDFASGEGNATWAPSLVQINPDITTVLARATNOGDWVDDSSRDNMLGVKKEGAAN 649  
Db 761 NOLEGEFAHADVDQKQILQONQADITNTATQONRNTVANGFEIEKNKAGI---ATN 816  
QY 650 SAALPGEMKPDVWSLNRSPGLG---LKPQPVQPKIDQPDQ--QPSQNTKN-VTPG 702  
Db 817 KQ-----ELIQNDRLNQINETNHQ--DOKIDOLGYALKEQOQHNNRISAV 862  
QY 703 NGDKPAGKATPDNTIDPSAQSGQNTIDPSAQXSGQNTKNV-TPGNEKQKQNTDAKOL 761  
Db 863 ERTAGIANATAIATPLSPSRAGEHHVLFVSGYHNGQAAVSLGAAGLSDTGKST----- 917  
QY 762 POTGNKSGLAGLYAGSL 778  
Db 918 ---YKIGLSWSDAGGL 930

RESULT 15  
US-08-409-995-2  
; Sequence 2, Application US/08409995  
; Patent No. 5646259  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen I.  
; APPLICANT: St. Geme III, Joseph W.  
; TITLE OF INVENTION: Haemophilus Adhesion Proteins  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Flehr, Hobbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/409,995  
; FILING DATE: 24-MAR-1995  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Silva, Robin M.  
; REGISTRATION NUMBER: 38,304  
; REFERENCE/DOCKET NUMBER: A-61053/RFT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1098 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
US-08-409-995-2

Query Match 4.1%; Score 170; DB 1; Length 1098;  
Best Local Similarity 20.3%; Pred. No. 0.00014;  
Matches 178; Conservative 100; Mismatches 301; Indels 300; Gaps 40;  
QY 22 GWTAHADVESNNYNGVAEYNTYRQANGQIGVDGKIIISANSNTTSGSTN-OESSATNNT 80  
Db 357 GLVTA-KVDIVAVNKTGWRIKTTD--ANGONG-DFAVATSGNTVTFASGNTTATVNGT 412  
QY 81 ENAVVNESKNTNTENAVVYNNENKNTNTENAVVNNENKNTNTENDNSQLKLTNNQPSAA 140

Db 413 DGITVYDAKVG---DGLKLDGDKIAADTTALTVDNCKNANPKGVADVASTDEKKLVTA 470  
QY 141 ---TOANLKKLNQAAKAVQNAKIDAGSLTDDQINELNKNINFSKAEGAKLTFKDLG 196  
Db 471 KGLVTAALNSLSMTTAAEA-----DCGTL-DGNASE-----QEVKAGDKVTFRAGKN 516  
QY 197 IGNAIVKQDPQVAPYSNAKEIKNNPATYVDAQTKMAHLVDWDSHPVQDPVTVGVSNY 256  
Db 517 L-----KVKQEGANET-----YSLQDALTLGLTSIT 541  
QY 257 MGQYLVIAMGIPNSPTGDNHIIYLLNYKYG---DNDFSHWRNAGSIFGTKEINVPQWS 312  
Db 542 LG-----TGNGAKTEINKDGLTITPANGAGANNANTISVTKG---ISAG 584  
QY 313 GSAIVNDGCTIQLF-----FTSNDTSYKLNQDRLATATLNLNVDDNGVSIKS---VD 362  
Db 585 GOSVKNVSGLKKFGDANPDLTSSADNLTQNDDAYKGLT---NLDEKGTDKQTPWAD 641  
QY 363 NYQVLFEGD--GFHYQTYEQFANGKDRENDYCLDRPHVVQLE-----NGDRY 408  
Db 642 NTAATV-GDLRGLGWVISADKTTGGSTEYHQ--VRNANEVFKSGNGINVSQKTVNGRRE 699  
QY 409 LVFEANTGTEDYQSDQIYNWANYGGDDAFNIKSSFKLLNNKKDRELAGLALGILKL 468  
Db 700 ITFELAKG-EVYKSNE-----FTVKET-----NGKETSIV----- 728  
QY 469 TNNQSKPKVEEYVSLVSTLMACDEVXXKGLDKYLLFSVTRYSRGSDRELTAKDNTIVGD 528  
Db 729 -----KVGDKYYSKEDIDITTGOPK---LKD-----GN 753  
QY 529 NVAMIGYVSDSLMGKYKPLNNSGVVLTASVPANWRTATYSYAVPVAGHPDQVLITSYMS 588  
Db 754 TVA-----AKYQDKGKGVSVTDNTEATITNKSGY-----VTGNQ-----VADATA 795  
QY 589 NKDFASGEGNATWAPSLVQINPDITTVLARAT-----NOGDWVDDSSRDNMLGVLK 644  
Db 796 KSGFELGLADEADAKRAF-----DDKTRALSAGTTEIVNAHDKVRFANGLNTRKVSAAATV 849  
QY 645 EGA-ANSAALPGEMKPDVWSLNRSPGLG-----KHPQPVQPKIDQPDQ 689  
Db 850 ESTDAN-----GDKVTTTFVKTDELPLTOIYNTDANGKKITKVVKDGQTKKYLNA 901  
QY 690 QPSGONTKNVTPGCDKPKAGKATPDN-----TN 717  
Db 902 DGTADMTREVTIGNVDSGKKVVDKNDGKWHAKADGTADKTGGEVSNDKYSTDEKHVVS 961  
QY 718 IDPSAQPSQONTNIDPSA-----OXSGONTKNVTPGNEKQ 752  
Db 962 LDPNDQSKGKGVWIDNVANGDISATSTDAINGSQLYAYAKGYTNLAGQ--VNNLEGKVNKV 1020  
QY 753 GKNTD-----AKQLPQTKNKGSLAGLYAGSLALFG 783  
Db 1021 GKRADACTASALAASQLPQ-----ATMPGKSMVAIAG 1052

Search completed: September 26, 2002, 18:24:27  
Job time: 46 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2002, 18:26:15 ; Search time 61.92 Seconds  
(without alignments)  
1420.711 Million cell updates/sec

Title: US-09-995-587A-11  
Perfect score: 4149  
Sequence: 1 MYKVGKNAVATVLSASILM.....LYAGSLALFLGLAAIEKRHA 792

Scoring table:  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_032802.\*  
1: /SIDSL/gcgdata/hold-geneseg/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSL/gcgdata/hold-geneseg/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSL/gcgdata/hold-geneseg/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSL/gcgdata/hold-geneseg/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSL/gcgdata/hold-geneseg/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSL/gcgdata/hold-geneseg/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSL/gcgdata/hold-geneseg/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSL/gcgdata/hold-geneseg/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSL/gcgdata/hold-geneseg/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSL/gcgdata/hold-geneseg/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSL/gcgdata/hold-geneseg/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSL/gcgdata/hold-geneseg/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSL/gcgdata/hold-geneseg/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSL/gcgdata/hold-geneseg/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSL/gcgdata/hold-geneseg/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSL/gcgdata/hold-geneseg/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSL/gcgdata/hold-geneseg/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSL/gcgdata/hold-geneseg/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSL/gcgdata/hold-geneseg/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSL/gcgdata/hold-geneseg/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSL/gcgdata/hold-geneseg/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSL/gcgdata/hold-geneseg/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	795.5	19.2	881	22	Novel human diagno
2	795.5	19.2	893	22	Novel human diagno
3	789.5	19.0	774	22	Novel human diagno
4	789.5	19.0	774	22	Novel human diagno
5	782	18.8	1027	22	Novel human diagno
6	773.5	18.6	736	22	Novel human diagno
7	762.5	18.4	789	22	Novel human diagno
8	762.5	18.4	789	22	Novel human diagno
9	760.5	18.3	527	22	Novel human diagno
10	760.5	18.3	789	22	Novel human diagno
11	760.5	18.3	823	22	Novel human diagno

12	758.5	18.3	493	22	ABG21647	Novel human diagno
13	758	18.3	546	22	ABG21832	Novel human diagno
14	758	18.3	562	22	ABG21605	Novel human diagno
15	758	18.3	1095	22	ABG21602	Novel human diagno
16	758	18.3	1095	22	ABG21823	Novel human diagno
17	755.5	18.2	855	22	ABG04500	Novel human diagno
18	752.5	18.1	929	22	ABG23199	Novel human diagno
19	751.5	18.1	472	12	AAR10671	B.amyloidoliquefacien
20	744	17.9	797	22	ABG18136	Novel human diagno
21	742.5	17.9	644	22	ABG06230	Novel human diagno
22	742.5	17.9	913	22	ABG02461	Novel human diagno
23	740.5	17.8	689	22	ABG00041	Novel human diagno
24	740	17.8	786	22	ABG02257	Novel human diagno
25	731	17.6	1551	22	ABG28409	Novel human diagno
26	723	17.4	542	22	ABG21613	Novel human diagno
27	690	16.6	657	22	ABG21612	Novel human diagno
28	690	16.6	657	22	ABG21834	Novel human diagno
29	668	16.1	914	22	ABG09624	Novel human diagno
30	668	16.1	3048	22	ABG25791	Novel human diagno
31	662.5	16.0	487	19	AAW48306	Bacillus sp. V230
32	654	15.8	778	22	ABG13422	Novel human diagno
33	651.5	15.7	828	22	ABG24549	Novel human diagno
34	651.5	15.7	928	22	ABG21821	Novel human diagno
35	651.5	15.7	928	22	ABG22394	Novel human diagno
36	651.5	15.7	928	22	ABG28921	Novel human diagno
37	646.5	15.6	436	22	ABG04298	Novel human diagno
38	646	15.6	495	22	ABG04320	Novel human diagno
39	643.5	15.5	708	22	ABG25779	Novel human diagno
40	643.5	15.5	708	22	ABG28401	Novel human diagno
41	634.5	15.3	692	22	ABG21599	Novel human diagno
42	633.5	15.3	893	22	ABG26977	Novel human diagno
43	622.5	15.0	426	22	ABG04288	Novel human diagno
44	595.5	14.4	981	22	ABG21856	Novel human diagno
45	586	14.1	417	22	ABG04289	Novel human diagno

## ALIGNMENTS

RESULT 1  
ABG21651  
ID ABG21651 standard; Protein; 881 AA.  
XX  
AC ABG21651;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #21642.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
XX  
N-PSDB; AAS85838.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess



```

Db 307 IANVTSIANSESINNGDTSIGPEDGSASEITRAFDTLAKALNTDSSSPSIADGID 366
QY 80 TENA-----VYNESEKNTN--NTENAVVNNKNTNTNENAVVNNKN-----TNTTNDNSQ 128
Db 367 TAGGSIHVSRDQSTPIIEVEGPILSDTHVTFKS-----IREDNGRSKQTVHTEGDMM 422
QY 129 LKLTNNEOPSAAQTANLKKLPQAQKAVONAKIDAGSLTDDQINELNKNFNSKSAEKGA 188
Db 423 -----NIKKIVKQATVLTFTTALLAGGATQAFAKENNGKAY-KETYGVSH 466
QY 189 LTFKDLLEGIGNAIVKODPOYATIPYNAKEIKNMPATYTVDAOTGKMAHLVDVDSWSPVQDP 248
Db 467 ITRHMLQIPKG--DQNEKYVPPQDQSTIKN-----LESAG-----LDVWDSWPLQ- 513
QY 249 VTGYVSNYMGYQLVIAMGIPNSPTGDHNYLLYNYKYGDNDFSHWRNAGSIFGTRETNYF 308
Db 514 ADGTVAEYNGYHVVALAGSPKD-ADDTSIYMFYQKVGDSIDSWKNAGRVF--KDSKDF 570
QY 309 -----QWSGSAIVNDGRIQIUFFTSNDTSYKLNQRLATATLNLNVDDNGVS 357
Db 571 DANDPILKDTQGEWSATFTSDGKIRIFYT--DYSGKHGKQSIITAGVNVSKSDTLK 628
QY 358 IKSVDNYQVLFEGDGFHYQTYEQFAN-GKDRENDYDCLRDPHVVOLENGDRYLVEANTG 416
Db 629 INGVEDHKTIFDQDGTQYQNVQGFIDEGNYTSGDHTLRDPHYVE-DKGHKYLVFEANTG 687
QY 417 TED-YOSDDQIYNWANYGGDDAFNIKSSFKLNNKKDRELALANGALGILKLTNNQSKP 475
Db 688 TENGYGGEESLFNKAYYGGTTFKESQKLQGSAKKRD-AELANGALGIELNNDYT-- 744
QY 476 KYEEVYSPLVSTLMACDEV-----XXKLGDKYVLFVSVTRVSRGSDRELTAKNFTIVGDNYA 531
Db 745 -LKKVMKPLITNTVTDEIERANVFMNGKWYLF-----DSRGSKMTI-----DGINSDIY 796
QY 532 MGYVSDSLMGKYPKPLNNGSVVLTASVPANWRTATYSYAVPVAGHPDOVLITSYMSNKD 591
Db 797 MGYVSNLSITGYPKPLNKTGLVLMQGLDPNDVFTYSHFAVPPQA-KGNVVTISYMTNRG 855
QY 592 FASGCGNTATWAPSLVQINPDPTTTVLARATNQG 626
Db 856 FF--EDKKATFAPSFLMNIKNGKTSVKNLSILEQG 888

RESULT 3
ABG05590
ID ABG05590 standard; Protein; 774 AA.
XX
AC ABG05590;
XX
DT 13-FEB-2002 (first entry)
DE
DE Novel human diagnostic protein #581.
XX
KW Human; Chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
WI WPI; 2001-639362/73.

```

```

DR N-PSDB; AA569777.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 20; SEQ ID No 35949; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity.
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 774 AA;

```

```

Query Match 19.0%; Score 789.5; DB 22; Length 774;
Best Local Similarity 34.3%; Pred. No. 8.4e-42;
Matches 222; Conservative 100; Mismatches 220; Indels 105; Gaps 27;

QY 144 NLKKLPQAQKAVONAKIDAGSLTDDQINELNKNFNSKSAEKAKLTFTKDLLEGIGNAIVK 203
Db 4 NIKKIVKQATVLTFTTALLAGGATQAFAKENNGKAY-KETYGVSHTRHMLQIPKG-Q 60
QY 204 QDPQYATIPYNAKEIKNMPATYTVDAOTGKMAHLVDVDSWSPVQDPVTGYVSNYMGYQLVI 263
Db 61 QNEKYVPPQDQSTIKN-----IESAG-----LDVWDSWPLQN-ADGTVAEYNGYHVVF 109
QY 264 AMNGIPNSPTGDHNYLLYNYKYGDNDFSHWRNAGSIFGTRETNYF-----QEWS 312
Db 110 ALAGSPKD-ADDTSIYMFYQKVGDSIDSWKNAGRVF--KDSKDFDANDPILKDTQGEWS 166
QY 313 GSAIVNDGRIQIUFFTSNDTSYKLNQRLATATLNLNVDDNGVSIKSVNYQVLFEGDG 372
Db 167 GSATFTSDGKIRIFYT--DYSGKHGKQSIITAGVNVSKSDTLKINGVEDHKTIFDGDG 224
QY 373 FHYQTYEQFAN-GKDRENDYDCLRDPHVVOLENGDRYLVEANTGTED-YOSDDQIYNWA 430
Db 225 KTYQNVQGFIDEGNYTSGDHTLRDPHYVE-DKGHKYLVFEANTGTENGYGGEESLFNKA 283
QY 431 NYGDDDAFNIKSSFKLNNKKDRELALANGALGILKLTNNQSKPKVEEYVSPLVSTLMA 490
Db 284 YYGGGTNTFFRKESQKLQGSAKKRD-AELANGALGIELNNDYT---LKKVMKPLITNTV 339
QY 491 CDEV-----XXKLGDKYVLFVSVTRVSRGSDRELTAKNFTIVGDVNMATGYVSDSLMGKYP 546
Db 340 TDEIERANVFMNGKWYLF-----DSRGSKMTI-----DGINSDIYMLGYVSNLSITGYPK 392
QY 547 LNNSGVVLTASVPANWRTATYSYAVPVAGHPDOVLITSYMSNKDPSASGENTATWAPSF 606
Db 393 LNKTVLMQGLDPNDVFTYSHFAVPPQA-KGNVVTISYMTNRGFF--EDKKATFAPS 449
QY 607 LVQINPDPTTTVLARATNQGDVWDSRSSNMLG-----VLKEG-----AANSAALPG 655
Db 450 LMIKNGKTSVKNLSILEQGLTWLVQAKRAGLGGGSGTGVIRVRVRIELASTHALA- 508

```





RESULT	7	
ABG04946		
ID	ABG04946	standard; Protein; 789 AA.
XX		
AC	ABG04946;	
XX		
DT	13-FEB-2002	(first entry)
XX		
DE	Novel human diagnostic protein #4937.	
XX		
KW	Human; Chromosome mapping; gene mapping; forensic;	
KW	food supplement; medical imaging; diagnostic; genetic disorder.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200175067-A2.	
XX		
PD	11-OCT-2001.	
XX		



Db 611 sqklqsakrd-aelangelgilelnndyt---lkkvmpkplitsntvdeieranvfkm 666  
QY 499 GKYLYLFSVTRVSRGSDRELTAKTNTIVGDNVAMIGYVSDSLMGKYKPLNNSGVLTASV 558  
Db 667 ngkwylyft---dsrgskmti---dginsndiymlygvnsnltgpykplnktglvlqmg1 719  
QY 559 PANWRATYSYAVPVAGHPDQVLITSYMSNKDFASGEGNYATWAPSLVQINPDPTTV 618  
Db 720 dpndvftysfhavpqa-kgnnnvitsymtnrgff--edkktafapsflmnikgnktsv 776  
QY 619 LARATNOQ 626  
Db 777 knsileqg 784

RESULT 8

ID ABG28407 standard; Protein; 789 AA.  
AC ABG28407;  
XX  
DT 18-FEB-2002 (first entry)  
DE Novel human diagnostic protein #28398.  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
OS Homo sapiens.  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSEQ-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS92594.  
XX

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -

Claim 20; SEQ ID NO 58766; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed

CC Specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 789 AA;

Query Match 18.4%; Score 762.5; DB 22; Length 789;  
Best Local Similarity 36.3%; Pred. No. 4.5e-40;  
Matches 199; Conservative 95; Mismatches 183; Indels 71; Gaps 21;

QY 101 ENKNTNTENAVVNEKN-----TNTTENDNSOLKLTNNQPSAATQANLKKUNLPQAAKA 155  
Db 286 knvdlkgfepdsirednrgsrqktvhtgdmnm-----nikkivkqatv1 330  
QY 156 VQNAKIDAGSLTDDQINELNKNFNSAKGAKLTFKDLBGIGNAIVKDPQYAIYPSNA 215  
Db 331 tfttallaggatqafakennqkay-ketygvshitrhmlqipkq--qqnekyqvvpqfdq 387  
QY 216 KEIKNMPATYTVDAQTGKMAHLVDWSDWPQDPVTGVSYNMGYQLVIAMMGIPNSPTGD 275  
Db 388 stikn-----iesakg---ldvwdswplqn-adgtvaeyngyhvvfalagspkd-add 435  
QY 276 NHIYLLYKYGDNDFSHWRNAGSIFGKTETNVP-----QWSSGSAIVDNDGTIQ 324  
Db 436 tslymfyqkvgnsdnsidwknagrff--kdsdkfdandpilkdtqewsgsatftsdgkir 493  
QY 325 LFFTSNDTSDYKLNQDLQATATLNLNVNDNGVSIKSVNDYOVLFEGDGFHYQTYEOPAN- 383  
Db 494 lfyf--dysgkygkqsltaqvnsksddtIkingvedhktfdgdgktyqvngqfide 551  
QY 384 GKDRENDYCLRDPHVQLENGDRYLVFETANTGED-YQSDDOIYNWANYGGDDAFNIKS 442  
Db 552 gnytsdgnhtlrdphyve-dkghkylvfeantgtengyggeeslfnkayvggtnffrke 610  
QY 443 SPKLLNKKDRELALANGALGILKLTNNQSKPKVEEVSPLVSTLMACDEV---XXKL 498  
Db 611 sqklqsakrd-aelangelgilelnndyt---lkkvmpkplitsntvdeieranvfkm 666  
QY 499 GKYLYLFSVTRVSRGSDRELTAKTNTIVGDNVAMIGYVSDSLMGKYKPLNNSGVLTASV 558  
Db 667 ngkwylyft---dsrgskmti---dginsndiymlygvnsnltgpykplnktglvlqmg1 719  
QY 559 PANWRATYSYAVPVAGHPDQVLITSYMSNKDFASGEGNYATWAPSLVQINPDPTTV 618  
Db 720 dpndvftysfhavpqa-kgnnnvitsymtnrgff--edkktafapsflmnikgnktsv 776  
QY 619 LARATNOQ 626  
Db 777 knsileqg 784

RESULT 9

ABG04307  
ID ABG04307 standard; Protein; 527 AA.  
XX  
AC ABG04307;  
XX  
DT 13-FEB-2002 (first entry)  
DE Novel human diagnostic protein #4298.  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
OS Homo sapiens.  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.



```
SQ Sequence 789 AA;
Query Match 18.3%; Score 760.5; DB 22; Length 789;
Best Local Similarity 36.1%; Pred. No. 6e-40;
Matches 197; Conservative 95; Mismatches 178; Indels 75; Gaps 21;

QY 99 VNEKNKNTNTEAVVNEKNKNTNTENDNSQLKLTNNEQPSAATQANLKLNPQAAKAVON 158
Db 298 Iredn-----grsqktvhtegdnm-----nikkivkqatvltft 333
QY 159 AKIDAGSLTDQINELNKNINFSKSAEKGAKLTFKDLGIGNAIVKQDPQYAIPIYSNAKEI 218
Db 334 tallaggtatqafakennqkay-ketygvshitrhdmqlipkq--qgnekyvqpfdqsti 390
QY 219 KNMPATYTVDAQTGKMAHLVDWDSVPQDPVGYVSNYMGYQVLVIAMGIPNSPTGDNHI 278
Db 391 kn-----iesakg----ldvwdswplqn-adgtvaeingyhvvfalagspkd-addtsi 438
QY 279 YLLYKNGDNDFSHWRNAGSIFGKTETNVF-----QEWSSGSAIVNDGDTIQLFF 327
Db 439 ymfykvgdngsidswknagrvt--kdsdkfdandpilkdtqegwsgatftsdgkirify 496
QY 328 TSNDTSDYKLNDRQATATLNLNVDDNGSVSKSDVNYQVLFEGDGFHYQTYEQFAN-GRD 386
Db 497 t--dysgkhygkqslttaqvnvksdtdtkingvedhktifdgdktkyqnvqqfidegny 554
QY 387 RENDDYCLRDHPHVQLENGDRYLVEFANTGTED-YOSDDQIYNWANYGGDDAFNIKSSPK 445
Db 555 tsgdhtlrdphyve-dkghkylvfeantgtengyggeeslnkayygggtntffrkesqk 613
QY 446 LLNNKDRRELKAGLALGILKLTNNQSKPKVEEYVSLVSTLMACDEV----XXKLGDK 501
Db 614 lqqsakkrd-aelangaigilelnndyt---lkkvmpkplittntvtdeieranvfkmgk 669
QY 502 YLFSVTRVSRGSDRELTAKNITVGDNVAMIGYVSDSLMGYKPLNNSGVVLTASVPAN 561
Db 670 wylft---dsrgskmti----dginndiymlygvsnsitgpykplnktglvlqmgldpn 722
QY 562 WRTATYIYAVPVAGHPDOVLITSMKNDPAGSGGNATWAPSELVQINPDPTTVTLAR 621
Db 723 dvftyshfavpqa-kgnnvitsymtnrgff--edkktafapsflmniknktsvvks 779
QY 622 ATNOG 626
Db 780 lileq 784

RESULT 11
ABG25769
ID ABG25769 standard; Protein; 823 AA.
XX AC
AC ABG25769;
XX DT
DT 18-FEB-2002 (first entry)
XX DE
DE Novel human diagnostic protein #25760.
XX KW
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX OS
OS food supplement; medical imaging; diagnostic; genetic disorder.
XX PN
PN Homo sapiens.
XX PD
PD WO200175067-A2.
XX PF
PF 11-OCT-2001.
XX PR
PR 30-MAR-2001; 2001WO-0508631.
XX PR
PR 31-MAR-2000; 2000US-0540217.
XX PR
PR 23-AUG-2000; 2000US-0649167.
XX PA
PA (HYSE-) HYSEQ INC.
```

```
XX PI
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR N-PSDB; AA389956.
XX PT
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX PS
XX Claim 20; SEQ ID No 56128; 103pp; English.
XX CC
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG0010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX SQ
SQ Sequence 823 AA;

Query Match 18.3%; Score 760.5; DB 22; Length 823;
Best Local Similarity 36.1%; Pred. No. 6.3e-40;
Matches 197; Conservative 95; Mismatches 178; Indels 75; Gaps 21;

QY 99 VNEKNKNTNTEAVVNEKNKNTNTENDNSQLKLTNNEQPSAATQANLKLNPQAAKAVON 158
Db 332 Iredn-----grsqktvhtegdnm-----nikkivkqatvltft 367
QY 159 AKIDAGSLTDQINELNKNINFSKSAEKGAKLTFKDLGIGNAIVKQDPQYAIPIYSNAKEI 218
Db 368 tallaggtatqafakennqkay-ketygvshitrhdmqlipkq--qgnekyvqpfdqsti 424
QY 219 KNMPATYTVDAQTGKMAHLVDWDSVPQDPVGYVSNYMGYQVLVIAMGIPNSPTGDNHI 278
Db 425 kn-----iesakg----ldvwdswplqn-adgtvaeingyhvvfalagspkd-addtsi 472
QY 279 YLLYKNGDNDFSHWRNAGSIFGKTETNVF-----QEWSSGSAIVNDGDTIQLFF 327
Db 473 ymfykvgdngsidswknagrvt--kdsdkfdandpilkdtqegwsgatftsdgkirify 530
QY 328 TSNDTSDYKLNDRQATATLNLNVDDNGSVSKSDVNYQVLFEGDGFHYQTYEQFAN-GRD 386
Db 531 t--dysgkhygkqslttaqvnvksdtdtkingvedhktifdgdktkyqnvqqfidegny 588
QY 387 RENDDYCLRDHPHVQLENGDRYLVEFANTGTED-YOSDDQIYNWANYGGDDAFNIKSSPK 445
Db 589 tsgdhtlrdphyve-dkghkylvfeantgtengyggeeslnkayygggtntffrkesqk 647
QY 446 LLNNKDRRELKAGLALGILKLTNNQSKPKVEEYVSLVSTLMACDEV----XXKLGDK 501
Db 648 lqqsakkrd-aelangaigilelnndyt---lkkvmpkplittntvtdeieranvfkmgk 703
QY 502 YLFSVTRVSRGSDRELTAKNITVGDNVAMIGYVSDSLMGYKPLNNSGVVLTASVPAN 561
Db 704 wylft---dsrgskmti----dginndiymlygvsnsitgpykplnktglvlqmgldpn 756
```

QY 562 WRTATSYAVVAGHPDQVLTYSYMSKDFASGEGNATWAPSELVQINPDPTTVLAR 621  
 Db 757 dvftysrhavpqa-kgnnvvtisymtnrgff--edkktafapsflmmikgnktsvvkns 813

QY 622 ATNOG 626  
 Db 814 lileqg 818

RESULT 12  
 ABG21647  
 ID ABG21647 standard; Protein; 493 AA.

XX AC ABG21647;  
 XX DT 18-FEB-2002 (first entry)  
 XX DE Novel human diagnostic protein #21638.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX OS food supplement; medical imaging; diagnostic; genetic disorder.

XX PN WO200175067-A2.  
 XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX DR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI: 2001-639362/73.

XX DR N-PSDB; AAS85834.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 XX PT diagnostics, forensics, gene mapping, identification of mutations  
 XX PT responsible for genetic disorders or other traits and to assess  
 XX PT biodiversity

XX PS Claim 20; SEQ ID No 52006; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and  
 XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 XX CC and gene mapping, and in recombinant production of (II). The  
 XX CC polynucleotides are also used in diagnostics as expressed sequence tags  
 XX CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 XX CC to restore normal activity of (II) or to treat disease states involving  
 XX CC (II). (II) is useful for generating antibodies against it, detecting or  
 XX CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 XX CC a food supplement. (II) and its binding partners are useful in medical  
 XX CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 XX CC disorders involving aberrant protein expression or biological activity.  
 XX CC The polypeptide and polynucleotide sequences have applications in  
 XX CC diagnostics, forensics, gene mapping, identification of mutations  
 XX CC responsible for genetic disorders or other traits to assess biodiversity  
 XX CC and to produce other types of data and products dependent on DNA and  
 XX CC amino acid sequences. ABG0010-ABG30377 represent novel human  
 XX CC diagnostic amino acid sequences of the invention.

XX CC Note: The sequence data for this patent did not appear in the printed  
 XX CC specification, but was obtained in electronic format directly from WIPO  
 XX CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 493 AA;

Query Match 18.3%; Score 758.5; DB 22; Length 493;  
 Best Local Similarity 36.1%; Pred. No. 4.2e-40;  
 Matches 197; Conservative 94; Mismatches 179; Indels 75; Gaps 21;

QY 99 VVENKNTNTNNTAVVNNKNTNTTENDNSQLKLTNNEOPSAAQTANLKLPPQAAKAVON 158  
 Db 2 iredrn-----grsktvhtegdmnm-----nikkivkqatvlft 37

QY 159 AKTDAGSLTDDQINELNKNINFSKSAEKGAKLTFFKDLGIGNAIVKQDPOYAIPIYSNAKEI 218  
 Db 38 tallagatdafakennqkay-ketygvshtrhdmliqpkq--gqnekyyvpgfdqgti 94

QY 219 KNMPATYTVDAQTGKMAHLVDWDSWPQDPTGVVSNYQVLYIAMMGIPNSPTGDNHI 278  
 Db 95 kn-----iesakg----ldwdswpqpn-adgtvaeingyvhvfalagspkd-addtsi 142

QY 279 YLLYNYKGDNDFSHWRNAGSIFGTKEYNVF-----QEWSGSAIVNDDGTIQLFF 327  
 Db 143 ymfyqkvgdnsidswnagrvm--kdsdkfdandpilkdtqewsgsatftsdgkirify 200

QY 328 TSNDTSDYKLNDRQRLATATLNLNVDDNGVSTKSDNYOVLFEGDGFHYQTYEQFAN-CKD 386  
 Db 201 t--dysakhygkqsltaqvnvksdctikngvedhktifdgdgktyqnvqqfidegny 258

QY 387 RENDDYCLRDPVHVQLENGDRYLVFEANTGTED-YQSDDDIYNWANYGGDDAFNIKSSFK 445  
 Db 259 tsgdnhtlrdphyve-dkghkylvfeantgtengyqgeesifnkayygggtfnfrkesqk 317

QY 446 LLNNKKORELAGLANGALGILKLTNNOSKPKVEEYVSPVSTLMACDEV----XXKLGDK 501  
 Db 318 lqqsakkrd--aelangalgielnddyt---lkkvmkplitsntvtdeieranvfkmgk 373

QY 502 YLFSVTRVSRGSDRELTAKDNTIVGDNVAMIGVYSDSLMGKYPKLNNSGVVLTASVPAN 561  
 Db 374 wylft---dsrgskmtl----dginsndiymlgvyshtltpykpInktglvlgmgldpn 426

QY 562 WRTATSYAVVAGHPDQVLTYSYMSKDFASGEGNATWAPSELVQINPDPTTVLAR 621  
 Db 427 dvftysrhavpqa-kgnnvvtisymtnrgff--edkktafapsflmmikgnktsvvkns 483

QY 622 ATNOG 626

Db 484 lileqg 488

RESULT 13

ABG21832  
 ID ABG21832 standard; Protein; 546 AA.

XX AC ABG21832;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #21823.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX XX



QY 57 KIISANSNTTSGTNOESSATNTNENAVVNESKNTN-NTENAVVVENKNTNTNENAVVNE 115  
 Db 26 klvgtkalsttg----kalrtlptakvfslpnlidfkvpalsilprksiredn---gr 78  
 QY 116 NKNTNTENDNSOLKLTNNQPSAATQAMKLLNPOAKAVQNAKIDAGSLTDQINELN 175  
 Db 79 sktvthtegdmnm-----nikkivkqatvltfttallagatqafakenn 123  
 QY 176 KINFSAEKGAKLTFKDLEGIGNATVKDPOVAIPYSNAKETKNPNATVYDAQTGKMA 235  
 Db 124 qkay-ketygvshtrhdmqlpkq--qnekyvvpqfdqstikn-----iesakg--- 171  
 QY 236 HLDVWDSWPQDPVTGYVSNYMGYQVLIAMGIPNSPTGDNHIIYLLYNYKYGNDPFSHWRN 295  
 Db 172 -ldvwdswplqn-adgtvaeyngyhvvfalagspkd-addtsiymfyqkvgsdswkn 228  
 QY 296 AGSIFGKTETNVF-----QEWSGSAIVNDGDTIQLPFTSDTSDYKLNLDQRLAT 344  
 Db 229 agrvf--kdsdkfdandpilkddtqewsgsatftsdkgrlrfyt--dysgkhygksltt 284  
 QY 345 ATNLNVDNDGVSISKVDNYQVLFEGDGFHYQYEQFAN-GKDRENDYCLDRPHVVQLE 403  
 Db 285 agnvksdtdtklnrvdthktifdgdktyqnvqfidegnytsdgnhtlrdphve-d 343  
 QY 404 NGDRYLVEANTGTED-YQSDQDIYNWANYGGDDAFNIKSFLLNKKDRELALANGA 462  
 Db 344 kghkylvfeantgtengyqgeeslfnkayvggntfnfrkesqlqqsakkrd-aelanga 402  
 QY 463 LGILKLTNNQSKPKVEEYVSPVSTLMACDEV---XXKLGDKYVLFVSFVTRVRSRGSREL 518  
 Db 403 lgielndndyt---lkkvmkpltsntvtdeieranvfkmgkwyift---dsrgskmti 456  
 QY 519 TAKDNTIVGDNVAMIGYVSDSLMGKXKPLNNSGVYLFASVPANWRTATYSYAVPVAGHP 578  
 Db 457 ----dginsndiymlygvsnsiltgpykplnktglvlqmgldpndvfttyshfavpqa-kg 511  
 QY 579 DOVLITSYMSKNDPASEGNYATWAPSLVQINPDPTTV 618  
 Db 512 nnvvtasymtnrgif--edkktafapsflmnikgnktsv 549

RESULT 15

ABG21602  
 ID ABG21602 standard; Protein; 1095 AA.  
 XX AC ABG21602;  
 XX DT 18-FEB-2002 (first entry)  
 XX DE Novel human diagnostic protein #21593.  
 XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX OS Homo sapiens.  
 XX PN WO200175067-A2.  
 XX PD 11-OCT-2001.  
 XX PF 30-MAR-2001; 2001WO-US08631.  
 XX PR 31-MAR-2000; 2000US-0540217.  
 XX PR 23-AUG-2000; 2000US-0649167.  
 XX PA (HYSB-) HYSEQ INC.  
 XX PI Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 XX DR N-PSDB; AAS85789.

PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 XX Claim 20; SEQ ID No 51961; 103pp; English.  
 PS The invention relates to isolated polynucleotide (I) and  
 XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 1095 AA;

Query Match 18.3%; score 758; DB 22; Length 1095;

Best Local Similarity 34.7%; Pred. No. 1.3e-39;

Matches 201; Conservative 106; Mismatches 199; Indels 74; Gaps 23;

QY 57 KIISANSNTTSGTNOESSATNTNENAVVNESKNTN-NTENAVVVENKNTNTNENAVVNE 115  
 Db 559 klvgtkalsttg----kalrtlptakvfslpnlidfkvpalsilprksiredn---gr 611  
 QY 116 NKNTNTENDNSOLKLTNNQPSAATQAMKLLNPOAKAVQNAKIDAGSLTDQINELN 175  
 Db 612 sktvthtegdmnm-----nikkivkqatvltfttallagatqafakenn 656  
 QY 176 KINFSAEKGAKLTFKDLEGIGNATVKDPOVAIPYSNAKETKNPNATVYDAQTGKMA 235  
 Db 657 qkay-ketygvshtrhdmqlpkq--qnekyvvpqfdqstikn-----iesakg--- 704  
 QY 236 HLDVWDSWPQDPVTGYVSNYMGYQVLIAMGIPNSPTGDNHIIYLLYNYKYGNDPFSHWRN 295  
 Db 705 -ldvwdswplqn-adgtvaeyngyhvvfalagspkd-addtsiymfyqkvgsdswkn 761  
 QY 296 AGSIFGKTETNVF-----QEWSGSAIVNDGDTIQLPFTSDTSDYKLNLDQRLAT 344  
 Db 762 agrvf--kdsdkfdandpilkddtqewsgsatftsdkgrlrfyt--dysgkhygksltt 817  
 QY 345 ATNLNVDNDGVSISKVDNYQVLFEGDGFHYQYEQFAN-GKDRENDYCLDRPHVVQLE 403  
 Db 818 aqnvksdtdtklnrvdthktifdgdktyqnvqfidegnytsdgnhtlrdphve-d 876  
 QY 404 NGDRYLVEANTGTED-YQSDQDIYNWANYGGDDAFNIKSFLLNKKDRELALANGA 462  
 Db 877 kghkylvfeantgtengyqgeeslfnkayvggntfnfrkesqlqqsakkrd-aelanga 935  
 QY 463 LGILKLTNNQSKPKVEEYVSPVSTLMACDEV---XXKLGDKYVLFVSFVTRVRSRGSREL 518  
 Db 936 lgielndndyt---lkkvmkpltsntvtdeieranvfkmgkwyift---dsrgskmti 989  
 QY 519 TAKDNTIVGDNVAMIGYVSDSLMGKXKPLNNSGVYLFASVPANWRTATYSYAVPVAGHP 578  
 Db 990 ----dginsndiymlygvsnsiltgpykplnktglvlqmgldpndvfttyshfavpqa-kg 1044  
 QY 579 DOVLITSYMSKNDPASEGNYATWAPSLVQINPDPTTV 618

Db 1045 nnvitsyntnrgff--edkktafapsflmnikgnktsvv 1082  
: |:|||||: |: | |:|||||: |: |: |

Search completed: September 26, 2002, 18:26:20  
Job time: 159 sec







Db 27 QNSPQKVPQFNASAIKNIDSAGYD-KSGLIDLVDWSWPLON-ADGTAANYGHYIV 84  
QY 263 IAMGIPNSPTGDNHLYLYLNKVGDNDFSHRNAGSIFGTETNV-----FQWSG 313  
Db 85 SALAGDPKN-SDTPUHLFYQKVGDTSIDSWKNAGRVFEDMDKFPVNDPYLYKYOTQWMSG 143  
QY 314 SAIVNDGDTIQLFFTSNDTSYKLN-----DORLATATLNLVNDNG-VSIRKVDN 363  
Db 144 SATLTKDGQVRLEYT-----DYSNPEDEGDTGAGNOIISTAQVNLQSDPAATLKVQCVSD 198  
QY 364 QVLPFE-GDGFHYQTYEQFAN-GKRENDYDCLDRPHVVOLENGDRYLVPFANTGTED-Y 420  
Db 199 HRSVFDGGDTYVQNIQQFIDEGKTSWGNHTLRDPHYVE-DKGRHLYVFEANTGTIDGY 257  
QY 421 QSDQIYNWANYGGDDAFNIKSSFKLLNNKKDBELAGLALGALGILKLTNNQSKPKVEEV 480  
Db 258 QDQSFNNKAYYGGSDVFFONEKNKLQSPK-KQIASLANGALGIVELADDDYT---VKSV 313  
QY 481 YSPLVSTLMACDEV---XXKLGDKYLFVSFVTRVSRGSDRELTAQNTIVGDNVAMIGYV 536  
Db 314 MKPLVASNTVADEVERANIFKNKNNKYLFTDSRGSKWTSDGINDKD-----VYMLPGP 366  
QY 537 SDSLCKYKPLNNSGVVLSPVAPNRRATYSYAVPVAGHPD--QVLITSYMSNKF 592  
Db 367 GDSLNGPHNPINETGLVLNNLDPADLTHTYSHCGIP---HPGNNVVLTSYMTNRGF 421

## RESULT 2

US-09-503-172A-2  
; Sequence 2, Application US/09503172A  
; Patent No. 6284510  
; GENERAL INFORMATION:  
; APPLICANT: ITO, Tetsuya  
; APPLICANT: FUJITA, Koki  
; APPLICANT: HARA, Kozo  
; APPLICANT: TONOUZUKA, Takashi  
; APPLICANT: SAKANO, Yoshiyuki  
; TITLE OF INVENTION: BETA-FRUCTOFURANOSIDASE GENE  
; FILE REFERENCE: 10749-0001-0  
; CURRENT APPLICATION NUMBER: US/09/503,172A  
; CURRENT FILING DATE: 2000-02-14  
; PRIOR APPLICATION NUMBER: JP 160416/1999  
; PRIOR FILING DATE: 1999-06-08  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; TYPE: PRT  
; ORGANISM: Arthrobacter sp.  
US-09-503-172A-2

Query Match 6.4%; Score 267; DB 4; Length 578;  
Best Local Similarity 23.8%; Pred. No. 1.2e-12;  
Matches 124; Conservative 82; Mismatches 196; Indels 118; Gaps 26;  
QY 200 ATVKODPOYALPYPSNAKEIKNNPATYTV-----DAOTGRKAHLVDWDSWPQDPVTGYVS 254  
Db 67 ADARLQMSDPTAPSR-NSMPASVTMTPTVPQDFPMSNEQVWVWDTWPLTDEDANQYS 125  
QY 255 NYMGQYLVIAAMGIPNSPTGDNHLY-----LLYNNKYG-----DNDFSHRNAGSIFGPK 303  
Db 126 -VNGWEIIFSLVADNLGDFDRHFAKICYFYRPAGVPAERPN--GWTYGLLVFKEG 182  
QY 304 ET-NVFO-----EWGSAIVNDGDTIQLFFT-----SNDTSDYKLNQRLATATLN 348  
Db 183 VTGQIFEDQSFQTSQWSSARVSKNGEIKLFFTDVAFYRNSDGTNIKPYDPRIALSVCK 242  
QY 349 LNVDDNGYSIKSVDNVQVLFEDGDGHYQTYEQFANGKRENDYDCLDRPHVVO--LENGD 406  
Db 243 VKANKKGVTLGFKNVKTLQADGHTYQTAGO-----NEFFNFRDPFTFEDPAHPGE 294  
QY 407 RYLVFPEANTGTEDYQSDQIYNWANYG---GD-----DAFNKSSFKLLNNKKDBELAG 457

Db 295 TMVFEGNSA---MORETATCNEADLGYRQGDPAETVDDVNASGATYQIGN----- 343  
QY 458 LANGALGILKLTNNQSKPKVEEVSPLYSTLMACD-----EVXXKLGDKYLYFSVTR--- 509  
Db 344 ----VGLAKAKNKQLTE--WEFLPPLISANCVTDQTERPOIYFKDG-KSYLFTLSHRGT 395  
QY 510 VSRGSDRELTAQNTIVGDNVAMIGYVSDSLMGKYKPLN-NSGVVLTASVPANW----- 562  
Db 396 FAAGLD-----GPEGVYGFVGDGIRSDYQPLNGSGGLALGNPTNLNLFGLGQPF 443  
QY 563 -----RTATSYAYVPVAGHPDQVLITSYMSNKRDFASGEGNYATWAPSLVQI--N 611  
Db 444 APDFNOHPGHFOAYSHYVMP--GGLVQSFIDTIGTHDDFVRG-----GTLAPTVMKDIGV 497  
QY 612 PDDTTTTLARA-TNQGDDWVDDSSRNDNMLGLVKEGANS 650  
Db 498 GDPTKTAVDYSYSEGLGGHADIPANKHLFTNGKFGVAVS 537

## RESULT 3

US-08-381-936-2  
; Sequence 2, Application US/08381936  
; Patent No. 5792923  
; GENERAL INFORMATION:  
; APPLICANT: ROBER, Manuela  
; APPLICANT: GEIER, Gebhardt  
; APPLICANT: GEIDER, Klaus  
; APPLICANT: WILLMITZER, Lothar  
; TITLE OF INVENTION: DNA sequences which lead to the  
; TITLE OF INVENTION: formation of polyfructans (levans), plasmids containing  
; TITLE OF INVENTION: these sequences as well as a process for preparing  
; TITLE OF INVENTION: transgenic plants.  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen  
; STREET: 1180 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: US  
; ZIP: 10036-8403  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/381,936  
; FILING DATE: 09-FEB-1995  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 93 02110  
; FILING DATE: 09-AUG-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE P42270618  
; FILING DATE: 08-DEC-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meilman, Edward A.  
; REGISTRATION NUMBER: 24,735  
; REFERENCE/DOCKET NUMBER: P/951-108  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 382-0700  
; TELEFAX: (212) 382-0888  
; TELEX: 236925  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 415 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-381-936-2

```
Query Match      5.7% Score 236.5; DB 1; Length 415;
Best Local Similarity 23.4%; Pred. No. 1.7e-10;
Matches 96; Conservative 67; Mismatches 152; Indels 95; Gaps 20;

QY 239 VDSHPVQDPVTGVSNMVGQVLIAMGIPNSPT-----GDHNI 278
Db 44 IWDTPMLRD-FDGEIISVNGWCIIITLTADRTNDPQFDENGNYDITRDWEDRHRARI 102
QY 279 YLLNKKYGDNDFSHRNAGSIFGTETNVFOEWSGSAI-VNDGDTIQLFFTSNDTSYKL 337
Db 103 CYWISRTGKD---WIFGRVMAEGVAPTTRWAGTPIILLNDRGDIDLYTCTVTPG--- 154
QY 338 NDQRLATATLNLNVDNDGVSISKSDNYQVLFEGDGFHYQYEQFANGKDRDNDYCLRDP 397
Db 155 --ATIAKVRGKIVTSDQSVSLEGFQOVTSLFSADGTIYQTEQ-----NAFWNFRDP 204
QY 398 H-VVQLENGDRYLVEANT-----GTEDYQSDQIYNWANYGGDDAFNIKSKFKLLNNKKD 452
Db 205 SPFIDRNDGKLYMLFEGNVAGPRGSHEITQ-----AEMG-----NVPPEGYEDVGGAKY 252
QY 453 RELAGLANGALGILKLTNNQSKPKVEEYVSPVSTLMACDEVXXK-----LGDKYILFSVT 508
Db 253 QA-----GCVG-LAVAKDLGSE-WQILPLITAVGVNDQTERPHFVFDGKYILFTIS 304
QY 509 RVSRGSDRELTAKDNTIVGDNVAMIGYVSDSLMGKYKPLNNSGVVLTASVPANWRTATYS 568
Db 305 -----HKYTFADNLTPGDGV--YGFVSDKLTGPTPMNSSGLVL--GNPSSQPFQYTS 353
QY 569 YVAVPVAGHPDQVLITSYMSN-----KDFASGEGNYATWAPSFVQINPD 613
Db 354 HYVMP-----NGLVTSFIDSVPWKGDYRIG---GTEAPTWKILLKGD 393

RESULT 4
US-08-943-374-2
; Sequence 2, Application US/08943374
; Patent No. 6028249
; GENERAL INFORMATION:
; APPLICANT: ROBER, Manuela
; APPLICANT: GEIER, Gebhardt
; APPLICANT: GEIDER, Klaus
; APPLICANT: WILLMITTER, Lothar
; TITLE OF INVENTION: DNA sequences which lead to the
; TITLE OF INVENTION: formation of polyfructans (levans), plasmids containing
; TITLE OF INVENTION: these sequences as well as a process for preparing
; TITLE OF INVENTION: transgenic plants.
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
; STREET: 1180 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10036-8403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,374
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/381,936
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P42270618
; FILING DATE: 08-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Mellman, Edward A.
; REGISTRATION NUMBER: 24,735
; REFERENCE/DOCKET NUMBER: P/951-108
```

```
TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 382-0700
; TELEFAX: (212) 382-0888
; TELEX: 236925
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 415 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-943-374-2

Query Match      5.7% Score 236.5; DB 3; Length 415;
Best Local Similarity 23.4%; Pred. No. 1.7e-10;
Matches 96; Conservative 67; Mismatches 152; Indels 95; Gaps 20;

QY 239 VDSHPVQDPVTGVSNMVGQVLIAMGIPNSPT-----GDHNI 278
Db 44 IWDTPMLRD-FDGEIISVNGWCIIITLTADRTNDPQFDENGNYDITRDWEDRHRARI 102
QY 279 YLLNKKYGDNDFSHRNAGSIFGTETNVFOEWSGSAI-VNDGDTIQLFFTSNDTSYKL 337
Db 103 CYWISRTGKD---WIFGRVMAEGVAPTTRWAGTPIILLNDRGDIDLYTCTVTPG--- 154
QY 338 NDQRLATATLNLNVDNDGVSISKSDNYQVLFEGDGFHYQYEQFANGKDRDNDYCLRDP 397
Db 155 --ATIAKVRGKIVTSDQSVSLEGFQOVTSLFSADGTIYQTEQ-----NAFWNFRDP 204
QY 398 H-VVQLENGDRYLVEANT-----GTEDYQSDQIYNWANYGGDDAFNIKSKFKLLNNKKD 452
Db 205 SPFIDRNDGKLYMLFEGNVAGPRGSHEITQ-----AEMG-----NVPPEGYEDVGGAKY 252
QY 453 RELAGLANGALGILKLTNNQSKPKVEEYVSPVSTLMACDEVXXK-----LGDKYILFSVT 508
Db 253 QA-----GCVG-LAVAKDLGSE-WQILPLITAVGVNDQTERPHFVFDGKYILFTIS 304
QY 509 RVSRGSDRELTAKDNTIVGDNVAMIGYVSDSLMGKYKPLNNSGVVLTASVPANWRTATYS 568
Db 305 -----HKYTFADNLTPGDGV--YGFVSDKLTGPTPMNSSGLVL--GNPSSQPFQYTS 353
QY 569 YVAVPVAGHPDQVLITSYMSN-----KDFASGEGNYATWAPSFVQINPD 613
Db 354 HYVMP-----NGLVTSFIDSVPWKGDYRIG---GTEAPTWKILLKGD 393

RESULT 5
US-08-362-232-2
; Sequence 2, Application US/08362232
; Patent No. 5641667
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Fructosyltransferase Enzyme, Method
; TITLE OF INVENTION: For its Production and DNA Encoding the Enzyme.
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann & Baron
; STREET: 350 Jericho Turnpike
; CITY: Jericho
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11758
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk, 3.5 inch, 1.44 Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 6.0 for DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,232
; FILING DATE: 22-December-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CU 125/93
; FILING DATE: 23-December-1993
```



```
Query Match      4.2%   Score 173; DB 3; Length 933;
Best Local Similarity 20.5%; Pred. No. 6.le-05;
Matches 162; Conservative 90; Mismatches 273; Indels 264; Gaps 40
```



; CURRENT FILING DATE: 1999-03-16  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 30  
; LENGTH: 1004  
; TYPE: PRT  
; ORGANISM: Haemophilus Influenzae  
US-09-268-347-30

Query Match 4.2%; Score 173; DB 4; Length 1004;  
Best Local Similarity 20.6%; Pred. No. 6.9e-05;  
Matches 191; Conservative 122; Mismatches 333; Indels 280; Gaps 46;

QY 5 GKNAVATL-----VSASILMGVVTAHADQVE-SNNYGV--AEYNTERRQANGQ 51  
Db 162 GKHTVTFLEKDLNVKATVSDKLSLG---ANGNKVDITSDTNGLKPAKSTNGQ-NGN 216  
QY 52 IGVDGKIISANSTSGSTNOESSATNTE-----NAVVNESK 89  
Db 217 VHLNG-TASTLDTITGTT---KSTAGVDVONHNRAASVADVLNAGNIOGNGASVDV 272  
QY 90 NTNNTENAVNENKNTNTNTENDNSQLKLTNEOPSAAATQANLKKLN 149  
Db 273 NTYDVTDFVNGLNTNVNVTDTAHNK-KTVRVVDVTGLPVQYVTEGTVVVKVNGEYEA 331  
QY 150 PQAQAVONAKIDAGSLTDDOI-----NELNKNFSKA-----ERGAULTFKDLEGIGNAI 201  
Db 332 KODGSADMDKKVENGKLAKTKVLVSANGTNPVKISNVADGTENTDAVSFKOLKAL----- 387  
QY 202 VKODPOVAIPYSNAKEIKNMPATVTVDAOTCKMAHLVDWDSWPVODPTGVYSNMGVOL 261  
Db 388 --QDKQVTLASNAY-----ANGSDADGK-----GIQT-----LSNGLNFKF 424  
QY 262 VIAMGIPNSPTGDNIHYLLYLNKYGDNDFSHWRNAGSIFGTETNVFOEWSGSAIVNDGD 321  
Db 425 -----KSTDGE-----LLNIKAEND-----TVTTPKGSVQVGGD 456  
QY 322 --TIQ-----LFTSNTSDYKLDQRLATATLNLNVDD-----NGVSTKS 360  
Db 457 KATIQDGAKTGTLVASELSVLNKLKVGKGTGCTGVTGDTHTDVLKSGDKVTLKA 516  
QY 361 VDNYQVLFEGDGFHYQYEOFANGKDRENDDYC-----LRDPHVQVLENGDRYLV 410  
Db 517 GDNLVKQEGTNETYALKDELTDVKSVEFKDTANGANGASTKITKDGLTITPANGAG--A 574  
QY 411 FEANTG-----TED-----YOSDDIYNWANYGGDAFN-IKSSFLLNNKDKRELAKLA 459  
Db 575 AGANTANTISVTKDGISAGNAKVNKVVSLGKFGDANFDPLTSSADNLTKQYDNAYKGLT 634  
QY 460 N-----GALGLK-----LNNQSKPKVEEYVSPVSTLMACDEVXX 496  
Db 635 NLDEKSKGKOTPTVADNTAATVGLRGLGWVISADTKGELNKEYNAQVRN---ANEVKE 691  
QY 497 KLGDKYLFVSIVSRSGSDRELT---AKDNTIVGDNVAMIGYVSDSLMGKYPKPLNNSGV 553  
Db 692 KSGNGINVSCKT-LDNGT-REITFELAKD-----ENAIAGSGSKALR-----DNTVAI 738  
QY 554 LTASVPANMRATYSYAYVPVAGHPDOVLITSYSNKK---DFASGEGNYATWAPSFVLQI 610  
Db 739 GTGNVNAEKSAGF-----GDP-----NYIEDRAGGSYAFGDNDRIT----- 775  
QY 611 NPDDTTIVLARATNOGDWVDDSDRNDMLG-----VLKGAANSALPGEWGKPVDMWSL 665  
Db 776 --SKNTEVLGNSVNA-----KRDANGNLVTEKEVYVKGDKGAKTKVTVPQALGETVENS 827  
QY 666 -----INRSPGLGLKPHQVPQKIDPPQPSGQNTKNTVPGNGDKPAGKATPDNTNI 718  
Db 828 YLGNASTATKDKKNLK-----SDGTAGNTTTAGATCTVNGFAGATAHGAHSV 875  
QY 719 DPSAQP-----SGQNTNIDPSAQ--XSGQNTKNVTPGNEKOGKNTD----- 757  
Db 876 GASGEERRIONVAAGEISATSTDAINGSQLYAYAKGVTNLAGQVKNYKGRADAGTASALA 935

QY 758 AKOLPOTGNKSLAGLYAGSLALFG 783  
Db 936 ASQLPQ-----ASMPGKSMVSIAG 954

RESULT 11  
US-08-446-855A-2  
; Sequence 2, Application US/08446855A  
; Patent No. 5849573  
; GENERAL INFORMATION:  
; APPLICANT: Stewart, Thomas S  
; APPLICANT: Flores, Maria V  
; APPLICANT: O'Sullivan, William J  
; TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl  
; TITLE OF INVENTION: phosphate synthetase II  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon & Vanderhye PC  
; STREET: 1100 No. 5849573th Glebe Road, 8th Floor  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22201-4714  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/446.855A  
; FILING DATE: 06-Jul-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mitchard, Leonard C  
; REGISTRATION NUMBER: 29.009  
; REFERENCE/DOCKET NUMBER: 47-80  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-816-4000  
; TELEFAX: 703-816-4100  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2391 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-446-855A-2

Query Match 4.2%; Score 173; DB 2; Length 2391;  
Best Local Similarity 20.3%; Pred. No. 0.00028;  
Matches 115; Conservative 93; Mismatches 227; Indels 132; Gaps 28;

QY 31 VESNNYNGVAEVNTERQANGQIGVDGKIISANSTSGST-----NOESSATNNTENAV 84  
Db 1275 VKNHLYNEVVD--DKDTQLH-----KENNNNNNNNGVKNCKLKNKESYGVNNSNCI 1326  
QY 85 VNESKNTN--NTENAVVNE--NKN---TNNTENAVV--NENKNTNT----- 122  
Db 1327 -----NTNNINENNICHDISINKNIKVTINNNSNISNNENVTNLCVSRAGSHHTY 1381  
QY 123 -----ENDNSQLKLTNNEOPSAAATQANLKKLPQAQAVONAKIDAGSLTDDQINEL 174  
Db 1382 GKEEKSIGSDDTNIIISAQNSNNFNNSCENNENKAN-----VDVNVLEND--TKKR 1429  
QY 175 KKNFSKSAEAGAKLTFKDLLEGIGNAIVKQDPQYAIPIYSNAKEIKNMPATYTVDAQTGM 234  
Db 1430 EDINTTVFMEGQNSVINNNKKNKENSLLKGDDEIVMVLKKNKNNYNSVINNVDCRCKDM 1489  
QY 235 AHLVDWDSWPVODPTGVYSN-VMGYQL---VIAMGIPNSPTGDNIHYLLYLNKYGDND 290  
Db 1490 -----DGKNINDECKTYKKKKYKDMGLNNNNIVDELNGTSHSTNDHLYLDNFTSDEEI 1543

QY 291 SHWRNAGSIFGKTKETNVFQEWSSAIVNDGDTIQLFFTSNDTSDYKLNQORLATATNLN 350  
DB 1544 GNNKNN-DWYLSKEKSIKNNPCNSYYVDSVY-----NNEYKINKMKELIDNENLN 1594  
QY 351 VDNQVSIKSDNYQVLFEGDGFHYQTYEQFANGKDR-----ENDDYCLRD-----PHVV 400  
DB 1595 -----DEYNNVNMNCSNYYNNAFAVNGKDRNDNLEND--CIEKNMDHYKHYN 1641  
QY 401 QLEN-----GDRYLVEFANTGTEDYQSDDOQIYNWANYGGDAPNFKSPFKLLNNKKORELA 456  
DB 1642 RNNRSTNRMMJVMNNEKESNHEKGHR--RNLGNKKKE-----KNMEKNKGNKKKNY 1696  
QY 457 GIANGALGILKLTNNQSPKVEEYSPVSTLMACDEVXXKLGDKYY-----LFSVTRVS 511  
DB 1697 HYVNH-----KRNEYNSNNTIESKFNYY-----DDINKK---EYEDENDIYYFTTHSS 1742  
QY 512 RGSDELTAQNTIVGDNVAMIGYVSD 538  
DB 1743 QGNNDL-SNDNYLSSEELNTDEYDD 1768

## RESULT 12

US-09-150-741-2  
; Sequence 2, Application US/09150741  
; Patent No. 6183996  
; GENERAL INFORMATION:  
; APPLICANT: Stewart et al.  
; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate  
; Patent No. 6183996  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/09/150.741  
; CURRENT FILING DATE: 1998-09-10  
; EARLIER APPLICATION NUMBER: PL6380  
; EARLIER FILING DATE: 1992-12-16  
; EARLIER APPLICATION NUMBER: AU93/00617  
; EARLIER FILING DATE: 1993-12-02  
; EARLIER APPLICATION NUMBER: 08/446,855  
; EARLIER FILING DATE: 1995-07-06  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 2391  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: protein  
US-09-150-741-2

Query Match 4.2%; Score 173; DB 4; Length 2391;  
Best Local Similarity 20.3%; Pred. No. 0.00028;  
Matches 115; Conservative 93; Mismatches 227; Indels 132; Gaps 28;  
QY 31 VESNNYNGVAEYNTERQANGQIGVDGKIISANSNTSGT-----NOESSATNNENAV 84  
DB 1275 VKNHLYNEVD-DKDTQLH-----KENNNNNNNNGVENCKLNKESGYNNSSNCI 1326  
QY 85 VNESKNTN--NTENAVNE---NKN-----TNTENAVY-NENKNTNNT----- 122  
DB 1327 -----NTNNINIENICHDISINKIKVTINNSNNSNENVENETNLNCVSRAGSHHIY 1381  
QY 123 -----ENDNSOLKLTNNQPSAATQALNKLNPQAQKAVQNAKIDAGSLTDDQINEL 174  
DB 1382 GKEKSGSDDTNLSRQNSNNFSCNENNNKAN-----VDVNVLEND-TKKR 1429  
QY 175 NKINFSKSAEKALTKFDELEGNAIVKQDPQYAIPIYSNAKEIKNNPATYTVDAQTGKM 234  
DB 1430 EDINTTVFMEGQNSVINNNKNSKSLKGDDEEDIVVNLKKNENYNNVNDCKKDM 1489  
QY 235 AHLVDWDSHPQDDPVTVGVSH-YMGYQL----VIAMMGIPNSPTGDNHYLLYLNKYGDNDF 290

DB 1490 -----DGKINDECKTYKKNKYKMDGLNNNNIYDELSNGTSHSTNDHLYLDNFTSDEEI 1543  
QY 291 SHWRNAGSIFGKTKETNVFQEWSSAIVNDGDTIQLFFTSNDTSDYKLNQORLATATNLN 350  
DB 1544 GNNKNN-DWYLSKEKSIKNNPCNSYYVDSVY-----NNEYKINKMKELIDNENLN 1594  
QY 351 VDNQVSIKSDNYQVLFEGDGFHYQTYEQFANGKDR-----ENDDYCLRD-----PHVV 400  
DB 1595 -----DEYNNVNMNCSNYYNNAFAVNGKDRNDNLEND--CIEKNMDHYKHYN 1641  
QY 401 QLEN-----GDRYLVEFANTGTEDYQSDDOQIYNWANYGGDAPNFKSPFKLLNNKKORELA 456  
DB 1642 RNNRSTNRMMJVMNNEKESNHEKGHR--RNLGNKKKE-----KNMEKNKGNKKKNY 1696  
QY 457 GIANGALGILKLTNNQSPKVEEYSPVSTLMACDEVXXKLGDKYY-----LFSVTRVS 511  
DB 1697 HYVNH-----KRNEYNSNNTIESKFNYY-----DDINKK---EYEDENDIYYFTTHSS 1742  
QY 512 RGSDELTAQNTIVGDNVAMIGYVSD 538  
DB 1743 QGNNDL-SNDNYLSSEELNTDEYDD 1768

## RESULT 13

US-08-169-927-2  
; Sequence 2, Application US/08169927  
; Patent No. 5783441  
; GENERAL INFORMATION:  
; APPLICANT: Carl, Mitchell  
; APPLICANT: Dobson, Michael E.  
; APPLICANT: Ching, Wei Mei  
; APPLICANT: Dasch, Gregory A  
; TITLE OF INVENTION: Gene and Protein Applicable to the  
; TITLE OF INVENTION: Preparation of Vaccines for Rickettsia prowazekii and  
; NUMBER OF INVENTIONS: Rickettsia typhi and the Detection of Both  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Counsel, Naval Medical R & D Command  
; STREET: Bldg. 1, T-12, 8901 Wisconsin Ave.  
; CITY: Bethesda  
; STATE: MD  
; COUNTRY: USA  
; ZIP: 20889-5606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/169,927  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/742,128  
; FILING DATE: 08/09/91  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Spevack, A. David  
; REGISTRATION NUMBER: 24,743  
; REFERENCE/DOCKET NUMBER: 75,976  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 295-6759  
; TELEFAX: (301) 295-1022  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1612 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-169-927-2

Query Match 4.1%; Score 172; DB 1; Length 1612;  
Best Local Similarity 19.5%; Pred. No. 0.00018;



Matches 200; Conservative 117; Mismatches 366; Indels 342; Gaps 47;

Qy	4	VGKWAATVLSASILMGVVTAHADQDVESNYNGVAEWNTER-----QANGOIG	53
Db	245	VSXKA-AFEFNVTGTGG-----NLKGIELNTAAVAGKLISLGAANAIVIG	291
Qy	54	VD-----CKIIS--ANSNTSG-----STNOESSAT--NNTENAV	84
Db	292	TDNGAGRAAGFIYVDNGNAATISGOVYAKNMVIOANAGGVTEHIVDVGLGTTNFK	351
Qy	85	VNESKTNNTNENAVERNKNNTN-----ENAVVNENKNNTNTE-----	123
Db	352	TADSK-----VIITENSFGSTNFGNLDITQIVVPDTKILKGNFIDGVKNNGTAGVIT	404
Qy	124	NDNSOLKLTNNQOPSAATOAN-----	144
Db	405	PNANGALYSASTDPNTAVTNIINAIEBAGVGVELSGIHIAELRONGGSIFKLADGTVIN	464
Qy	145	-----LKKLNPOAAKAVQ--NAKIDAGSLDDQIN-ELNKINFKSBAERKALTKFDL	194
Db	465	GPVNQNALMNNALAGSIQDGSAITIGDIGNGGVNAALQHITLANDASKILADGANI	524
Qy	195	EG--IGNAIVKDPQOYAIPYSNAKEIKNPARYTYDAQTGRMAHLDVDWSPVOD-PWTG	251
Db	525	IGANVGAIHFQANGTGIKLTNQ--NNIVNVFDLDTDKTVGVVDASSLTNNQTLTING	582
Qy	252	YVSVMGYQLVIAAMGIPASPT-----GD-----NHLYLLNRYKYGONDESHRNA	296
Db	583	SIGTVANTKTLAQNLIGSSKITLNAQDVAINELVIENNGSVOLHNHTYLITKTINAAQ	642
Qy	297	GSI-----PQTKETNV-----PQMSGSAIV	317
Db	643	GOIIIRAADPLNTWTTLDAGTNLGSAENPLSTHFATKAANADSILNVKGYNLVANNITT	702
Qy	318	NDOGTQLRFTSNDTS-----DYKLNDRLATATNLNVDDNCGVSKSDVNTQVL	367
Db	703	NDANVSLHFRSGCTSIVSGTVGGQGKLNLIIL-----DNQTTVK-----	744
Qy	368	FEGDGHFYQYEOFANGKDRENDXYCLRDHPHYVQLENG--DRYLVFEANTCETYQSDQ	425
Db	745	FLGD---TF--NGGTK-----IECKSLIQISNNYTTHDHVESADNTGLEFVNTDP	790
Qy	426	I-----YNWANYG-----GDAENIKSPKLLN---NKKDRELACLANG-----	461
Db	791	ITVTLNKQAYFOVLKVQTIISPGNIVFNEIGNVGIVHGIAANSISFENASLGTSFLPS	850
Qy	462	--ALGIKILTNNQSKPEBEVYSPLVSTILMACDEVXXK---LGDKYYLFSTRVSRGSDR	516
Db	851	GTPLDLVLTIKSTVNGTCVDNFENAPIV-VVSGIDSMINGQIIGDKK---NIALSLGSDN	906
Qy	517	ELTAKNTIVGDNVAMIGVVSLSMKGYKPLNNSG-VVLTASVPAN-----	561
Db	907	SIRVNANTLYS-----GIRTKNQCTVTLSGGMNPNPCITYGLGLENGSP	952
Qy	562	-WRTATY-----SYAPVPVAGHPDOVLITSYMSKDF-----ASCEGNYATWAP	604
Db	953	KLKQVTFETTDYNNLGSIIANNVTINDVDTLTLTGGIAGTDFDAKTILSVGNNAV---	1007
Qy	605	SFLVQNPDDTTTVLARATNOGD-----YWDDSSRRDNMLGYLKEAGANSALPGE-	656
Db	1008	RFXDSTFSDPBSRIVATQANKGTVTYLGNALYSNIGSLDTPVASVRFTGNDSGAGLOGNI	1067
Qy	657	WGKPVDW----SLINRSPLGLKHQPQPKIDOPDOQPSQONTKNTVTPGNGDKPAGKA	711
Db	1068	YSQNIDFGTYNLTILSNVILG-GGTTAINGEID-----LITNLLIFANGFTSWGDN	1118
Qy	712	TPDNTIDPSAOPSQ--NTNIDPSAQSGONTKNVTPGNEKQCKNTDAKQLPOTCNK-	767
Db	1119	TSISTLLNVSSNGIGOVTAEDAQVNAITTTGITIKIQDNANANFSGTQAYTLIQGARF	1178
Qy	768	SLAG 772	
Db	1179	NGTLG 1183	

RESULT 14  
US-09-336-447A-9  
; Sequence 9, Application US/09336447A  
; Patent No. 6310190  
; GENERAL INFORMATION:  
; APPLICANT: HANSEN, ERIC J.  
; APPLICANT: AEBI, CHRISTOPH  
; APPLICANT: COPE, LESLIE D.  
; APPLICANT: MACIVER, ISOBEL  
; APPLICANT: FISKE, MICHAEL J.  
; APPLICANT: FREDENBURG, ROSS A.  
; TITLE OF INVENTION: USPAl AND USPAl2 ANTIGENS OF MORAXELLA CATARRHALIS  
; FILE REFERENCE: AMCY:024  
; CURRENT APPLICATION NUMBER: US/09/336,447A  
; CURRENT FILING DATE: 1999-06-21  
; NUMBER OF SEQ ID NOS: 98  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 9  
; LENGTH: 941  
; TYPE: prt  
; ORGANISM: Moraxella catarrhalis  
US-09-336-447A-9

	Query Match	4.1%	Score 170.5;	DB 4:	Length 941;
	Best Local Similarity	20.7%;	Pred.	No. 9.8e-05;	Mismatches 190; Conservative 108; Mismatches 380; Indels 239; Gaps
QY	4	VGKNWAVATLVASILMGVVVTAHADOVESNNNGVAEVTNRQAQGICGVDG-----KI 58 : :   :			
Db	111	IGKNSTV-----GGGTNDA-----MGEYSTVAG--GANQAKGNSTVGCGGNKA 155 : :   :			
QY	59	ISANSNTTSGTSNQ-----ESSATTNTENAVVNESKTNTNTENAVERNKNNTN 106 : :   :			
Db	156	IGNNSTVGGSNQNAKGHSTIAGKKNNQTATGNSPAAGVENKADAN--NAVAGHNKNTI 213 : :   :			
QY	107	NTENAVNVNEKNT-----NNTENDNSQ---LKLTNREQPSAATQAANKLN-- 149 : :   :			
Db	214	EGTNSVAIGSNNTVTKTGKENPVFILGSNTNTENAQSGLVLCGNTTAGKAAATTVNAEVNL 273 : :   :			
QY	150	-----POAAKAVQN--AKIDAGSLTDQINELINKINFSK---SAEGKAKLTFFDLBGIGN 199 : :   :			
Db	274	TLENFAGASKANANNITGVSVGS----ENNERTIVNAGGISATSTDVAYNGSQLHALAK 329 : :   :			
QY	200	AIVRQDDPYAIPIYSNAKEIKNM-PATTVDAQTGKMA-----HLVDWSWPQDVPVGY 252 : :   :			
Db	330	AVAK-----NKSDIKGLNGVKELDKVEGVLSRDINSLUHDVADN---QDSIAKN 376 : :   :			
QY	253	VSNTMGYQLVIAMM-----GIPSPNTGDNHITLYLLNYKYGDNDFSHWRNAGSIGF----- 301 : :   :			
Db	377	KADIKGLNKEVKELDKEVGLSRDIGS-----LHDDVADNQDSIAKNKADIKGLNKEVKE 431 : :   :			
QY	302	-TRETNVFQEWSSSAIVNDGTTQLFTSNDTSDYLKDORLATATVLN---VDDONGVS 357 : :   :			
Db	432	LDEKVGLSRDIGS--LHDDVATNQADIANKQADIKTLNNNVEEELLNLSGRLLDQKADI 489 : :   :			
QY	358	IKSVDNYOVLFEFGGFHYTYEQFANG-----KRENDYCYLRDPHVQ 401 : :   :			
Db	490	DNNINNITYEAQQODQHSSDIKTLKNVBEGLLDLSRLDQKADITKDTIKTLKN-----N 545 : :   :			
QY	402	LENG-----DRIFYEANTGTEDYQSDQI---YNWANYGDDDAFNIAIK 441 : :   :			
Db	546	VEEGLLDLSRLDQKADIANKQADIAQNQOTDIQDLAAYNELQDOYAQKQTEAIDLAKA 605 : :   :			
QY	442	SS-----FKLLNNKKRELAGLANCGLIKLTTNNQPKPVEEYVPLVSTLMAC 491 : :   :			
Db	606	SSANTDRATAELGIAENKKDAQIA-----KAQANENKDGJ-----AK 643 : :   :			
QY	492	DEVXXKLGDKY-----YLFsvTRvRGSDRELtAKDNTlvgdNWAmig----- 534 : :   :			
Db	644	NQADIQIHDKRKTI NGLTGHSMVARVCNNIT QGVATNKADITAKNQADIANNKNIYELAQQ 703 : :   :			

QY 535 ---YVSD-SLMGKYPLNNSGVLT-ASVPANWRATYSYAVPVAGHPDQVLTISYMSN 589  
 Db 704 QDQSSDIKTAKYSAANTDRIAKNKAEDASFETLTKNQNTLIEQG---EALVQONKAI 760  
 QY 590 KDFASGEGNYATWAPSFVLQNPDDTTVTLARATHOGDWDSSRDNDMLGVLKGAAN 649  
 Db 761 NOLEGFRAHADVDQKQILQONQADITTNKTALEQINIVTANGFEIEKNKAGI-----ATN 816  
 QY 650 SAALPGENGKPVMSLINRSPGLG-----LKPHPVQPKIDQDQ--QPSGQNTKN-VTPG 702  
 Db 817 KQ-----ELILQRLNLAQINETNNHQ--DQKIDQLGYALKKEQGGHFNRRISAV 862  
 QY 703 NGDKPAGKATPONTNIDPSAQSGONTNIDPSAQSGQNTKNV-TPGNEKQKNTDAQL 761  
 Db 863 ERQTAGGIANIAIATLPSRAGEHHVLFSGYHNGQAAVSLGAAGLSDTGKST----- 917  
 QY 762 PQTGNKSLAGLYAGSL 778  
 Db 918 ---YKIGLSWSDAGGL 930

## RESULT 15

US-08-409-995-2

; Sequence 2, Application US/08409995

; Patent No. 5646259

; GENERAL INFORMATION:

; APPLICANT: Barenkamp, Stephen I.

; APPLICANT: St. Geme III, Joseph W.

; TITLE OF INVENTION: Hemophilus Adhesion Proteins

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESS: Flehr, Hohbach, Test, Albritton &amp; Herbert

; STREET: Four Embarcadero Center, Suite 3400

; CITY: San Francisco

; STATE: CA

; COUNTRY: USA

; ZIP: 94111-4187

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/409,995

; FILING DATE: 24-MAR-1995

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Silva, Robin M.

; REGISTRATION NUMBER: 38,304

; REFERENCE/DOCKET NUMBER: A-61053/RFT

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 781-1989

; TELEFAX: (415) 398-3249

; TELEX: 910 277299

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1098 amino acids

; TYPE: amino acid

; STRANDEDNESS: double

; TOPOLOGY: unknown

US-08-409-995-2

## Query Match

Best Local Similarity 4.1%; Score 170; DB 1; Length 1098;

Matches 178; Conservative 100; Mismatches 301; Indels 300; Gaps 40;

QY 22 GVVTAHADQVSNYNGVAEVTERQACQIGVKGKLIISANSNTTSGSTN-QESSATNNT 80  
 Db 357 GLVTA-KDVDAVNTKGTWRIKTTD--ANGQNG-DFAIVASGNTVTFASGNGTATVNTGT 412  
 QY 81 ENAVNVESKNTNTNENAVNVENKNTNTNENAVNVENKNTNTNENDNSQLKLTNNEQPSAA 140

Db 413 DGITVKYDAKVG--DGLKLDGDKIAADTTALTVDNCKNANNPKGVADVASTDEKKLVTA 470  
 QY 141 ----TQANLKLNPQAARAVQNAKIDAGSLTDDQINELNKNINFSKSAEKGAKLTFKDLG 196  
 Db 471 KGLVLTALNSLSWTTAAEA-----DGGTL-DGNASE-----QEVKAGDKVTFKAGKN 516  
 QY 197 IGNAIVKQDPQVAIPYSNAKEIKNMPATTYVDAQTQKMAHLVDWDSWPVQDPVTGVSYNY 256  
 Db 517 L-----KVKQEGANFT-----YSLQDALTLGLTSIT 541  
 QY 257 MGYQLVIAMGIPNSPTGDNHIYLLYNKYG----DNDFSHWRNAGSIFGFKETNVFQWS 312  
 Db 542 LG-----TGNGAKTEINKDGLTITPANGAGANNANTISYTKDG---ISAG 584  
 QY 313 GSAIVNDGDTIQLF-----FTSNDTSYKLNQDLATATLNLNVDNDNGSVKIS---VD 362  
 Db 585 GGSVKNVVSGLKKFGDANFDPLTSSADNLTQNDDAYKGLT---NLDEKGTQKQTPVAD 641  
 QY 363 NYQVLFEGD--GFHYQTYEQFANGKDRDNDYCLRDPHVVVQLE-----NGDRY 408  
 Db 642 NTAATV-GDLRGLGWISADKTTGGSTGYHDQ--VRNANEYKFKSGNGINVSQKTVNGRRE 699  
 QY 409 LVFEANTGTEDYQSDDOIYNWANYGGDAFNKISFKLLNNKKDRELALAGLALGILKL 468  
 Db 700 ITFELAKG-EVVKSN-----FTVKET-----NGKETS LV----- 728  
 QY 469 TNNQSKPKVEEYVSPVSTLMACDEVXXKLGDKYYLFSTVTRVSRGSDRELTAKDNTIVGD 528  
 Db 729 -----KVGDKYYSKEDIDLTTGQPK---LKD---GN 753  
 QY 529 NYAMIGYVSDSLMGKYKPLNNSGVLTASVPANWRATYSYAVPVAGHPDQVLTISYMS 588  
 Db 754 TVA-----AKYQDKGGKVVSVTDNTEATITNKGSGY-----VTGNQ-----VADAIA 795  
 QY 589 NKDFASGEGNYATWAPSFVLQINPDQTTVLARAT-----NOGDWVWDDSSRDNDMLGVLK 644  
 Db 796 KSGFELGLADEADAKRAF-----DDKTRALSAGTTEIYNAHKVRFANGLTKVSAATV 849  
 QY 645 EGA-ANSAALPGEMGKPVDMWSLINRSPGLGL-----KPHQVPQPKIDQPDQ 689  
 Db 850 ESTDAN-----GDKVTTTFYKTDVLPVLTQIYNTDANGKKITKVYKDGQTKWYELNA 901  
 QY 690 QPSGQNTKNVTPGNGDKPAGKATPDN-----TN 717  
 Db 902 DGTADMTKEVTLGNVDSGKVVYKVDNDGKWHAKADGTADTKGCVSNDKVSTDEKHVVS 961  
 QY 718 IDPSAQPSQNTNIDPSA-----QXSGQNTKNVTPGNEKQ 752  
 Db 962 LDPNDQSKGKGVVIDNVANGDISATSDAINGSQLYAVAKGVNTLAGQ-VNNLEGGKVNK 1020  
 QY 753 GKNTD-----AKQLPQTGKNSGLAGLYAGSLLALFG 783  
 Db 1021 GKRADAGTASALAASQLPQ-----ATMPGKSMVAIAG 1052

Search completed: September 21, 2002, 23:37:43  
 Job time: 5084 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2002, 18:27:55 ; Search time 63.39 Seconds  
(without alignments)  
2161.413 Million cell updates/sec

Title: US-09-995-587A-11  
Perfect score: 4149  
Sequence: 1 MYKVGKNWAVTLVSASILM.....LYAGSLLALFGLAAIKRHA 792

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL\_19.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertibrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	685.5	16.5	489	16 Q97I81	Q97I81 clostridium
2	676	16.3	499	2 Q925E5	Q925E5 paenibacill
3	662.5	16.0	487	2 O82854	O82854 bacillus sp
4	315	7.6	428	16 Q97I79	Q97I79 clostridium
5	245.5	5.9	630	2 Q9EVD6	Q9EVD6 actinomycetes
6	238.5	5.7	424	2 Q93FD9	Q93FD9 pseudomonas
7	237.5	5.7	431	2 Q93TM1	Q93TM1 pseudomonas
8	235.5	5.7	415	2 Q9AEX9	Q9AEX9 rahnella aq
9	233	5.6	1395	2 Q9AISO	Q9AISO staphylococ
10	230	5.5	1171	2 Q9KWX6	Q9KWX6 staphylococ
11	219.5	5.3	622	5 Q9U0K4	Q9U0K4 plasmodium
12	212.5	5.1	930	5 Q97I51	Q97I51 cryptospori
13	207.5	5.0	1318	5 Q95PH4	Q95PH4 dictyosteli
14	203.5	4.9	2817	16 Q97K42	Q97K42 clostridium
15	199.5	4.8	1166	2 O86489	O86489 staphylococ
16	198.5	4.8	1599	2 Q00599	Q00599 streptococc

17	198	4.8	2276	2 Q93TY6	Q93TY6 staphylococ
18	195.5	4.7	1093	16 Q9CGV6	Q9CGV6 lactococcus
19	193.5	4.7	1141	16 Q99W46	Q99W46 staphylococ
20	193	4.7	629	16 Q92DS1	Q92DS1 listeria in
21	192.5	4.6	1141	16 Q932F7	Q932F7 staphylococ
22	191.5	4.6	1119	5 Q97374	Q97374 plasmodium
23	189	4.6	1125	16 Q98PM9	Q98PM9 mycoplasma
24	188.5	4.5	931	2 Q9K113	Q9K113 staphylococ
25	184	4.4	604	5 Q26021	Q26021 plasmodium
26	184	4.4	1463	2 O86919	O86919 staphylococ
27	183.5	4.4	3628	5 Q968Y0	Q968Y0 plasmodium
28	183	4.4	559	5 Q9U3Y8	Q9U3Y8 plasmodium
29	183	4.4	2706	5 Q97292	Q97292 plasmodium
30	183	4.4	3890	16 Q99U53	Q99U53 staphylococ
31	182.5	4.4	1650	5 Q77328	Q77328 plasmodium
32	182.5	4.4	3610	5 Q968T7	Q968T7 plasmodium
33	182.5	4.4	3705	2 Q9F285	Q9F285 versinia pe
34	182	4.4	935	16 Q932C5	Q932C5 staphylococ
35	182	4.4	989	16 Q99VJ4	Q99VJ4 staphylococ
36	182	4.4	1449	2 O68542	O68542 streptococc
37	182	4.4	4919	2 Q92HL0	Q92HL0 haemophilus
38	182	4.4	6713	16 Q99U54	Q99U54 staphylococ
39	182	4.4	6713	16 Q931R6	Q931R6 staphylococ
40	181.5	4.4	947	2 O86487	O86487 staphylococ
41	181	4.4	836	17 Q9HM69	Q9HM69 halobacteri
42	180.5	4.4	1529	5 Q9GQC2	Q9GQC2 dictyosteli
43	180.5	4.4	3620	5 Q968T6	Q968T6 plasmodium
44	180	4.3	732	2 Q9F0R0	Q9F0R0 staphylococ
45	179	4.3	1063	16 Q9CH86	Q9CH86 lactococcus

#### ALIGNMENTS

RESULT 1

Q97I81 PRELIMINARY; PRT; 489 AA.

AC Q97I81; DT 01-OCT-2001 (TREMBlrel. 18, Created)

DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE LEVANSUCRASE.

GN CAC1772

OS Clostridium acetobutylicum.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;

OC Clostridium.

OX NCBI\_TaxID=1488;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;

RX MEDLINE=21359325; PubMed=11466286;

RA Noelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.,

RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,

RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,

RA Bennett G.N., Koonin E.V., Smith D.R.

RT "Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum."

RL J. Bacteriol. 183:4823-4838(2001).

DR EMBL; AEO07686; AAK79737.1; -

DR InterPro; IPR003469; Glyco\_hydro\_68.

DR Pfam; PF02435; Glyco\_hydro\_68; 1.

KW Complete proteome.

SQ SEQUENCE 489 AA; 54553 MW; 9002CB364FD1D3CD8 CRC64;

Query Match 16.5%; Score 685.5; DB 16; Length 489;

Best Local Similarity 33.3%; Pred. No. 1.8e-26;

Matches 170; Conservative 101; Mismatches 167; Indels 69; Gaps 18;

QY 161 IDAGSLTDDQINELNKLNFKSAEKAGKATFKDLGIGN-----AIYKQDPQVAIP 211

DB 16 ICSGIITQTFASINDMN-----YKETYGVSHITRYNMSKIPMSNDLKKFVP 63

212	Qy	Y	SNAKEIKNMPATYTVDAOTGKMAHLDWDSWPQDPVTGYYSNTWGWYQVLVTAMGIPNS	271
64	Db	Q	FNASTLKIATSAKGYD-KNGNLIDLJDWDSWPLQNG-DGTVANHYGHVIFALAGDPKN	121
272	Qy	P	TGDNHIIYLLNKGNDFSHWRNAGSIEGTETNV-----FOEWSGSAIYVNDGOT	322
122	Db	Q	DDTSIYWFYQKIGNSIDSWKNAGKVFKSDSKYVANDPYLYKQTOEWSGSATVTSQOQ	180
323	Qy	I	QLFFT-----SNDTSDYKLNDQRLATATLNLN-VDDNGSISIKSYDNTQVILFE-GD	371
181	Db	V	RLFYDFDSGVAKDGGTDA\$-----NOVITTTQVNL\$QDP\$N\$TINIDSYS\$D\$K\$S\$V\$F\$D\$G\$N	235
372	Qy	G	FHVOYQYEOFAN-GKDRENDYCLRDPHVVOLENGDRYLVFEANTGCTD-YOSDDQIYNW	429
236	Db	G	TYIYNVQOQIDBGKSSGDNHFLRDPHVFE-DNGRKRYLVFEANTGTDNGYOGDTSLLNK	294
430	Qy	A	NYGGDDAFNIKSFFKLLNKKDKRELAGLANGALGTKLITNNQ\$K\$P\$K\$E\$V\$Y\$S\$P\$V\$T\$U\$M\$	489
295	Db	A	FYGRSQ\$F\$K\$T\$E\$K\$O\$Q\$U\$LL\$D\$T\$N\$K\$K\$H\$D\$A\$S\$L\$A\$N\$G\$A\$G\$I\$E\$L\$N\$D\$Y\$T\$--L\$K\$E\$M\$K\$P\$L\$A\$S\$T\$	351
490	Qy	A	CDEV-----XXKLGDKXYILF\$V\$T\$R\$V\$S\$G\$S\$D\$R\$E\$L\$T\$A\$K\$D\$N\$T\$I\$G\$D\$N\$V\$A\$M\$I\$G\$Y\$V\$S\$D\$S\$L\$M\$K\$Y\$K\$	545
352	Db	V	TEI\$E\$R\$A\$N\$V\$F\$K\$M\$G\$R\$W\$Y\$L\$F\$T\$S\$R\$G\$S\$K\$M\$T\$I\$N\$G\$S\$S\$K\$D\$-----I\$Y\$M\$G\$F\$S\$N\$S\$U\$T\$G\$P\$Y\$K\$	404
546	Qy	P	LAN\$G\$V\$V\$L\$T\$A\$S\$P\$A\$N\$W\$R\$T\$A\$S\$Y\$A\$V\$P\$A\$G\$H\$D\$Q\$V\$L\$T\$S\$Y\$M\$S\$N\$K\$D\$F\$A\$S\$G\$E\$N\$Y\$A\$T\$W\$A\$P\$S	605
405	Db	P	L\$A\$N\$T\$G\$L\$V\$L\$N\$L\$D\$P\$D\$L\$F\$T\$Y\$S\$H\$F\$A\$V\$P\$O\$T\$-N\$G\$K\$N\$V\$V\$T\$S\$Y\$T\$N\$R\$G\$W\$Y\$S\$-D\$H\$S\$S\$F\$A\$P\$S	461
606	Qy	F	LVQI\$N\$P\$D\$D\$T\$T\$V\$L\$A\$R\$N\$O\$G\$D\$W\$V\$D\$D	632
462	Db	F	LLN\$K\$T\$K\$T\$S\$V\$S\$N\$S\$I\$Q\$O\$Q\$U\$T\$D\$N	488

```

RESULT      2
Q925E5
ID          PRELIMINARY;      PRT;      499 AA.
AC.         Q925E5;
DT          01-MAY-1999 (TREMBLrel. 10, Created)
DT          01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT          01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE          LEVANSUCRASE.
DE          SABC.
GN          PAENIBACILLUS polymyxa (Bacillus polymyxa).
OS          Bacteria; Firmicutes; Bacillus/Clostridium group;
OC          Bacillus/Staphylococcus group; Paenibacillus.
OX          NCBI_TaxID=1406;
OC          [1]
RN          SEQUENCE FROM N.A.
RP          STRAIN=CF43;
RC          MEDLINE=21042000; PubMed=11200435;
RX          Bezzate S., Aymerich S., Chambert R., Czarnes S., Berge O., Heulin T.
RT          "Disruption of the Paenibacillus polymyxa levansucrase gene impairs
RT          its ability to aggregate soil in the wheat rhizosphere.";
RL          Environ. Microbiol. 2:333-342(2000).
DR          ENBL: AJ133737; CAB39327.1; -
DR          Interpro: IPR003459; Glyco_hydro_58.
DR          Pfam: PF02435; Glyco_hydro.68; 1.
SQ          SEQUENCE 499 AA; 55317 MW; E646CD986292336C CRC64;

```

Query Match	16.3%	Score 676;	DB 2;	Length 499;
Best Local Similarity	34.5%;	Pred. No. 5.3e-26;		
Matches 180;	Conservative	91;	Mismatches 172;	Indels 78;
				Gaps 22;

Qy	162	DASLTDD----	QINELNKINFSKSAEKAKLTFKOLEGIGNAIWKQDPQYAIPIYSNAKE	217
Db	33	DASLTKDRAFTQITRNDMLNFKQOKG	-----BKYEVPSEDAST	72
Qy	218	IKNMPATYTYDAQTGKMAHLDDWDSRPVDPQVPTGYVSNYAGYQLVIAMGIPNSPTGDNH	277	
Db	73	IKNPISAMGRDS--SGKLIDPDWWSWPLQN-ADGTYVANKGNIVIFGLAGDPKRGE-DTF	129	
Qy	278	IYLLNKYKGNDFSHWRNAGSIEGTKE-----TNYFQEWMSGSAIVYNDGCTIGLQFT	328	

Db	130	IYLFYQAGNTSLSGWKAGRVKDNNDKLLANDPILKNQSEWGSATLTSDSQVLFY	189
Qy	329	SNDRSDYKLN---DORLATLNLN-VDDNGVSTKSDYDNQVLF--GDGHFYQTYEQFAN	383
Db	190	SROQYDPNNNLYGKQTLSTAQINVSQPDQTKLDIGVEDLKSIVDGGDKIYNVQQ--SV	248
Qy	384	GKRENDYCLDRPHVVOLENGORYLVFEANTGTE-DYQSDDOIYNWANYGGDADFNIKS	442
Db	249	GVDMDN---HTFRDPHYVE-DQGHKYIIFEANTGTETGYQGEDSIQNYPYVGGNKKEFTEE	305
Qy	443	SFKLLNNKRELAGLANGALGILKLTNNOSKPKVEEYSPLYSTLMACDEV----XXKL	498
Db	306	QONLLQSPK-KKGELANGALGIVELNDDYT---LKNVMPPLASNLYTDEIRANVFKM	361
Qy	499	GDKYLFSTVRSGSDRELTKADNTITGVDNVAMIGYVSDSLMGKYKPLNNSGVLTASV	558
Db	362	NGLWYLFSTRGSK-----VTDAIGDDDIYMLGYVSTLTGPYKPLNGTGLVLHQDL	414
Qy	559	PANWRTATSYIYAPVAGHPDQVLITYSYMSKNDFASGEGVNYAWFASFLVOINPDDTTTV	618
Db	415	DRDDVTYTAHFAIP-CKGNVNVVSSYMTNRGLF--PDHKSFTAFSLNIKGSYTSVV	471
Qy	619	LARATNGDWYDDSSRNDNMLGVLEKGAANSALPGEWCK	659
Db	472	KNGLIEGOITVDPT--NDK-----KES-----PNEYGK	498

RESULT	3		
ID	082854	PRELIMINARY;	PRT; 487 AA.
AC	082854;		
DT	01-NOV-1998	(TrEMBLrel. 08, Created)	
DT	01-NOV-1998	(TrEMBLrel. 08, Last sequence update)	
DT	01-JUN-2001	(TrEMBLrel. 17, Last annotation update)	
DE	BETA-FRUCTOFURANOSIDASE.		
OS	Bacillus sp.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;		
OC	Bacillus/Staphylococcus group; Bacillus.		
OX	NCBI_TaxID=1409;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-V230;		
RA	Tsusaki K.;		
RL	Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-V230;		
RA	Kurimoto M.;		
RL	"Cloning and sequencing of beta-fructofuranosidase gene from <i>Bacillus</i>		
RT	sp. V230."		
RL	Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AB010272; BAA32083.1; -		
DR	InterPro; IPR003469; Glyco_hydro_68.		
DR	Pfam; PF02435; Glyco_hydro_68; 1.		
SQ	SEQUENCE 487 AA; 53412 MW; 7F52A47921824AE3 CRC64;		

Query Match	Score	DB 2:	Length
Best Local Similarity	16.0%	662.5;	487;
Matches 168;	35.3%;	pred. No. 2.4e-25;	
Conservative	84;	Mismatches 141;	Indels 83; Gaps 19;

[illegible]

QY 316 IVNDGDTQLFPTSDTSDYKLN-----DQRLATATLNLVDDNG-VSIKSVNDYQ 365  
 Db 178 TUKDQGVRLFT-----DYSNPDGDTGAGNQLIISTAVNLSQPDAAATLKVDGVSDBK 232  
 QY 366 VLFE-GDGFHYQTYRQFAN-GKDREDDYCLDRPHVVOLENGDRYLVEFANTGTED-YQS 422  
 Db 233 SVFDGDDGTVYQNIQFIDEGKWSGDNHRLDRPHYVE-DKGHKYLVEFANTGTGDTGQ 291  
 QY 423 DDQIYNWANYGGDDAFNTSKSEKLLNNKKRELALANGALGILKLTNNQSKPKVEEYVS 482  
 Db 292 DQSFNNKAYYGGSDVFFQNEKNKLQSPK-KOIASLANSALGALGIVELADDT---VKSVMK 347  
 QY 483 PLVSTLMACDEV---XKLGDKYVLFVSVTRVSRGSDRELTAQNTIYVDGNVAMIGYSD 538  
 Db 348 PLVASNTVADEVERANIFPMKNKWTFLFDSRGSKMTSGINDKO-----YMLGPGGD 400  
 QY 539 SLMGKYKPLNNSGVVLTASVPANNRTATSYAVPVAGHPD--QVLITSYMSNKKDF 592  
 Db 401 SLNGPHNPINETGLVNLNLPADLTHTYSHGIP---HPGNNVVLTSYMTNRGF 453  
 RESULT 4  
 Q97179 ID Q97179 PRELIMINARY; PRT; 428 AA.  
 AC Q97179;  
 DT 01-OCT-2001 (Tremblrel. 18, Created)  
 DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE LEVANSUCRASE.  
 GN CAC1774.  
 OS Clostridium acetobutylicum.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 ON NCBI\_TaxID=1488;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;  
 RX MEDLINE-21359325; PubMed-11466286;  
 RA Neelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,  
 RA Gibson R., Lee H.M., Dubois J., Ghu D., Hitti J., Wolf Y.I.,  
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,  
 RA Bennett G.N., Koonin E.V., Smith D.R.;  
 RT "Genome sequence and comparative analysis of the solvent-producing  
 bacterium Clostridium acetobutylicum";  
 RL J. Bacteriol. 183:4823-4838(2001).  
 DR EMBL; AE007686; AAK79739.1; -  
 DR InterPro; IPR003469; Glyco\_hydro\_68.  
 DR Pfam; PF02435; Glyco\_hydro\_68; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 428 AA; 48395 MW; 9670B154B178E23E CRC64;

Query Match 7.69; Score 315; DB 16; Length 428;  
 Best Local Similarity 27.18; Pred. No. 3e-08;  
 Matches 105; Conservative 68; Mismatches 158; Indels 56; Gaps 18;  
 QY 236 HLDVWDSHPQDPTGYVSNYMGVQLVIAMGIPNSPTGDNHIIYLLNKKYGNDFSHRN 295  
 Db 66 NLVWMDTWPLVKK-DGSLAVVNGYKVFALPASRNVGNKRDHVGISYFCSTGDENVY 124  
 QY 296 AGSIFGKETNVFQEWGSAIVNDGDTQLFPTSDTSDYKLN--DQRLATATLNLVDD 353  
 Db 125 KGLAYNVEDALGSQWAGSAILDENGWQFFYTATGRKGEAVRFEQRLVTKFSINVDK 184  
 QY 354 NGVSKSVNDYQVLFEGDGFHYQTYRQFANGKREDDYCLDRPHVVOLENGDR-YLVE 412  
 Db 185 GGVHITNCSKHQVILEPDGVYQTMQO-AKGP-----IIYSRDPYFFEDPKTKDYLF 239  
 QY 413 ANTGTE-DYQSDDDQIYNWANYGGDDAFNKSFKLLNNKKRELALANGALGIL----- 466  
 Db 240 GNKGKIEKMKPENI-----GDKLFR-----KNHIAIPRGVENF-NGNVGIAVQNK 284

QY 467 KLTNNSKPKVEEYVSPVSTLMACDEVXXXKLGDKYVLFVSVTR---VSRGSDRELTAQDN 523  
 Db 285 DLTRFKLLPLLEAVG--VNOQLERQPIVMK-KNKYILEFTISHKFTYAQG-----LNGVD- 336  
 QY 524 TIVGDNVAMIGVSDSLMGKYKPLNNSGVVLTASVPANNRTATSYAVPVAGHPDQVLI 583  
 Db 337 -----GLYFCGNSLRNYPKPLNGNLVITN--PTNDPYQTSWYL--VSGHVLVSFI 385  
 QY 584 TSYMSNKKDFASGEGNYATWAPSLVQI 610  
 Db 386 NEYHFNGQLRYG---GTFAPTLQISL 408  
 RESULT 5  
 Q9EVD6 ID Q9EVD6 PRELIMINARY; PRT; 630 AA.  
 AC Q9EVD6;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE FRUCTOSYLTRANSFERASE.  
 GN FTF.  
 OS Actinomyces naeslundii.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Actinomycetaceae; Actinomycetaceae; Actinomycetes.  
 ON NCBI\_TaxID=1655;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 12104, WVU45;  
 RT Bergeron L.J., Morou-Bermudez E., Burne R.A.;  
 RT "Characterization of the fructosyltransferase gene of Actinomyces  
 naeslundii";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF228582; AAG09737.1; -  
 DR InterPro; IPR003469; Glyco\_hydro\_68.  
 DR Pfam; PF02435; Glyco\_hydro\_68; 1.  
 KW Transferase.  
 SQ SEQUENCE 630 AA; 68256 MW; CDCA04184DD6D7D2 CRC64;

Query Match 5.99; Score 245.5; DB 2; Length 630;  
 Best Local Similarity 22.7%; Pred. No. 0.00013;  
 Matches 150; Conservative 91; Mismatches 285; Indels 135; Gaps 31;  
 QY 75 SATNTENAVNENKNTNTENAVNENKNTNTENAVNENKNTNTENDNSQLKLTNN 134  
 Db 26 AGTGTALADETPSPPTSSATAEASPEAGDQKNDQSPADQQAQTPTDQGAQADQ 85  
 QY 135 EQP--SAATQAN-LKKLNPOAAKAVONAKIDAGSLTDDQINELNKINFSKSAEKAKLTF 191  
 Db 86 QAAGGAADODANGFRADNPGWANATKHTGAAHG-----VEENYTAKTADAMQIQRVF 139  
 QY 192 KLEGIGNAIVKQDPQYALPYSSNAKEIKN-MPATYTVDAQTGKMAHLVDVWDSWPDQVPT 250  
 Db 140 QP-QGAVRAPTRLPQLTMP-----EISNGFPAT-----SEDVWVWDTWTLDEAA 184  
 QY 251 GYVSNYMGVQLVIAMGIPNS--PTGDNHIY-----LLYKNYGDN----DFSHRNAGSIF 300  
 Db 185 HOIS-YNGWEIAFSLVADRAGHYTFDDRTHARLGFYRKAGTQTSADGAGSNGGWY 243  
 QY 301 G-----TKETNVF--QEWGSAIVNDGDTQLFPTS-----NUTSDY----- 335  
 Db 244 GGHVFPDASRSRTRVHAQTEMSGARLMENKIRMFYTSVPSTTAPRTGTTTHQV 303  
 QY 336 ----KLNDQRLATATLNLVDDNGVSIKSVNDYQVLFEGDGFHYQTYRQFANGKREDDY 392  
 Db 304 ATSKYDPRIVQSEGRIYATKNGVLTGFTQHQLLVPGKYYQFREQ-----NPGV 355  
 QY 393 CLURDPHVVOLENG-----DRYLVEFANTGTEDYQSDDDQIYNWANYGGDDAFNKSFKLLN 448  
 Db 356 NFRDPFTFRDQNNPDPTEYMYFEGNSA---FVREQYVDAAAKAGQNTTLATCTEEDLG 412  
 QY 449 NKK----DRELALANGALGILKLTN-----NQSCKPVE-EVYSPLVSTLMACDEVXXX -- 497

```
Db 413 YEKGPRAETVEANQGGYQLANGLARAKNKAMTEWEYLPPLLSGNCVNDQTERPQI 472
Qy 498 --LGDKYLFVSVRSGSDRELTAQNTIVGDNVAMIGYVSDSLMGKYKPLN-NSGWL 554
Db 473 YFODGKYLFETI-----SHRETYA--DGLQGPB-GYVGFVGDLRSDYKPLNQTGIAL 523
Qy 555 TASVPANWRTA-----TYSYAVPVACHPDQVLITYSNKNDFASGGNY 599
Db 524 GNPILNMFNPKYSPDFNSPYTFQSYHYVMP--GGLVESEFDSIGGNKGNPNVRG-- 579
Qy 600 ATWAPSLVQINPDPTTVILARATNOGDWDDSSRRNDMLGVLKEGAANSALPGEWK 659
Db 580 GSLSPVTKLNIS-GDTFSV-----DRTYCTNGLGGFADIPADRARTNGATR 625
Qy 660 P 660
Db 626 P 626

RESULT 6
ID Q93FU9 PRELIMINARY; PRT; 424 AA.
AC Q93FU9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE LEVANSUCRASE.
GN LSCA.
OS Pseudomonas aurantiaca.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=86192;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S-4380;
RA Koh I., Jang E.K., Kim J.S., Song K.S., Kim C.H., Rhee S.K.;
RT "Screening, molecular cloning and characterization of a novel
RT levansucrase, and its expression in Escherichia coli.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF306513; AAL09386.1;
SQ SEQUENCE 424 AA; 47009 MW; 9C1B4936A985162E CRC64;

Query Match 5.7%; Score 238.5; DB 2; Length 424;
Best Local Similarity 22.9%; Pred. No. 0.00018;
Matches 95; Conservative 76; Mismatches 152; Indels 91; Gaps 21;

Qy 239 VVDSWPVQDPVTGYVSNYMGYQLVIAMWG--IPNSPT-----GDNH-----I 278
Db 51 INDTMPLRD--LOGNITSVDGWSVIFLTADRHHPNDPEYIDENGNYDITRDWDRHGRAM 109
Qy 279 YLLNYKGNDFSHRNAGSIFGKTETNVFQWGSASAI--VNDGDTIQLFFT--SNDTSDY 335
Db 110 YVWFSRTGRD---WKLGRVMAEGVSPARWAGTPIILLNCGVDLYTAVTFCATIV 165
Qy 336 KLNDQRLATATLNLNVDDNGVSIKSDVNVQVLFEGDGFHYQYEQFANGKRENDYCLR 395
Db 166 KYRGRVVT-----EHGVSNVGFVKPLFEADGKMYQTEAQ--NPFWGFDPWFER 215
Qy 396 DPHVQLENGDRYLVEANTGTEDYQSDDOIYNWANYGDDAFNKSFKLLNKKREL 455
Db 216 DPN-----DGKLYMLFEGNVAGE--RGSHKV-----GKAEIGDVPVPGYEDVGNR---- 258
Qy 456 AGLANGALGILKLTNNQSKPKVEEYVSPVLSVTLMAACDEVXXK----LGDKYVLFVS 511
Db 259 --FOTACVGIARVEDGDGD--WEMLPPLLTAAGVNDQTERPHFVFGQKYLFTISHT- 313
Qy 512 RGSDELTAQNTIVGDNVAMIGYVSDSLMGKYKPLNNSGVVLTASVPANWRTATYSYA 571
Db 314 -----FTYADGVTPGDGV--YGFVADSLFGPYVPLNGSGLVL--GNPSSQPFQYSHCV 363
Qy 572 VPVAGHPDQVLITSYM-----SNKDFASGEGNATWAPSLVQINPDITTV 618
```

```
Db 364 MP-----NGLVTSFIDSVPIDESTGTOIRIG-----GTEAPTVKIKGQQTFFV 407

RESULT 7
ID Q93TM1 PRELIMINARY; PRT; 431 AA.
AC Q93TM1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE LEVANSUCRASE LSCC (EC 2.4.1.10).
GN LSCC.
OS Pseudomonas syringae (pv. glycinia).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=318;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PG4180;
RX MEDLINE=98394981; PubMed=9726857;
RA Hettwer U., Jacekel F.R., Boch J., Meyer M., Rudolph K., Ullrich M.S.;
RT "Cloning, nucleotide sequence, and expression in Escherichia coli of
RT levansucrase genes from the plant pathogens Pseudomonas syringae pv.
RT glycinia and P. syringae pv. phaseolicola.";
RL Appl. Environ. Microbiol. 64:3180-3187(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=PG4180;
RX MEDLINE=21242726; PubMed=11344135;
RA Li H., Ullrich M.S.;
RT "Characterization and Mutational Analysis of Three Allelic lsc Genes
RT Encoding Levansucrase in Pseudomonas syringae.";
RL J. Bacteriol. 183:3282-3292(2001).
DR EMBL; AF346402; AAK4952.1;
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 431 AA; 47619 MW; 0C71D0F133071AA5 CRC64;

Query Match 5.7%; Score 237.5; DB 2; Length 431;
Best Local Similarity 23.2%; Pred. No. 0.0002;
Matches 95; Conservative 66; Mismatches 154; Indels 95; Gaps 18;

Qy 239 VVDSWPVQDPVTGYVSNYMGYQLVIAMGIPNSPTGDNH-----IYLLYN----- 283
Db 60 INDTMPLRE--LOGTVVSVNGWSVILT-----TADRHPPDQYLDANGRYDKRDWED 111
Qy 284 KYGDNDFSHRN-----AGSIFGKTETNVFQWGSASAI--VNDGDTIQLFTSNDTSD 334
Db 112 RHGRARMCYWYSRTGKDWIFGGRVMAEGVSPTRWAGTPIILLNKGDDILTYTCVTPG- 170
Qy 335 YKLNQRLATATLNLNVDDNGVSIKSDVNVQVLFEGDGFHYQYEQFANGKRENDYCL 394
Db 171 -----AAVAKVRGRIVTSQGVKELDKFTQVKLFADGTYQT-----EAQSSWNF 217
Qy 395 RDPH--VVOLENGDRYLVEANTGTEDYQSDDOIYNWANYGDDAFNKSFKLLNKKDR 453
Db 218 RDPSPFIIDNDGKLYWFEVGNV-----AGERSHTVGAELGPVPPGH 260
Qy 454 ELAANGALGILKLTNNQSKPKVE--EYVSPVLSVTLMAACDEVXXK----LGDKYVLFVS 508
Db 261 EDVGARGFQVGCIGLAVAKDLSGEWEILPPLVTAAGVNDQTERPHYVFGDQKYLFTIS 320
Qy 509 RYRSRSDRELTAQNTIVGDNVAMIGYVSDSLMGKYKPLNNSGVVLTASVPANWRTATYS 568
Db 321 -----HKFTYADGVTGPDGV--YGFVGEHLFGPYRPMNASGLVL--GNPPEQPFQY 369
Qy 569 YYAVVAGHPDQVLITSYM-----KDFASGEGNATWAPSLVQINPD 613
Db 370 HCVMP-----NGLVTSFIDSVPIDESTGTOIRIG-----GTEAPTVRILLKGD 409

RESULT 8
```







```

QY 470 NNQSKPKVEEVYSPVSTL 488
DB 518 NNSLLKKKEEQINEYINEI 536

RESULT 12
077151
ID 077151 PRELIMINARY; PRT; 930 AA.
AC 077151
DT 01-NOV-1998 (TremBLrel. 08, Created)
DT 01-NOV-1998 (TremBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE HC-23 PROTEIN.
GN HC-23.
OS Cryptosporidium parvum.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Elmeriida;
OC Cryptosporidiidae; Cryptosporidium.
OX NCBI_TaxID=5807;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98398455; PubMed=9729442;
RA Schroeder A.A., Brown A.M., Abrahamson M.S.;
RT "Identification and cloning of a developmentally regulated
RT Cryptosporidium parvum gene by differential mRNA display PCR.";
RL Gene 216:327-334(1998).
DR EMBL; AF076438; AAC95010.1; -
DR InterPro; IPR000571; Zf-CCCH.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00642; Zf-CCCH; 1.
DR SMART; SM00184; RING; 1.
DR SMART; SM00356; Znf_C3H1; 1.
SQ SEQUENCE 930 AA; 98563 MW; 5AD78C87FB61F190 CRC64;

Query Match 5.1%; Score 212.5; DB 5; Length 930;
Best Local Similarity 20.8%; Pred. No. 0.0088;
Matches 174; Conservative 99; Mismatches 307; Indels 255; Gaps 39;

QY 21 GGVTAHQVNESNYNGVAEVNTERQANGQIGVDGKIISANTSTSGSTN-----Q 72
DB 212 GGVLN---DGFNNHSHNSVALQNN-----TVNHTTSGSTNGSFISDPWM 253

QY 73 ESSATNTNENAVVNE-SKNTNTNENAVVNEKNNTNTE-----NAVVNEKNNTNTENDN 126
DB 254 STSGYSPNGFLHDENSYDNNNSNLPNMNCNTEGISTDTNTNINHMKCQNGANG- 312

QY 127 SOLKLTNNQPSAATQANLKLNPQAKAVQNAKIDAGSLTDDQINELNKFNSKA--E 184
DB 313 ----ILNNGQNSKIVNGAMNLS-----ETRNSIETDKKYCDLINEGKNKDLXQPGINV 363

QY 185 KGAKLTDFKLEGNAIVKQDPQVAIPYSNAKEIKNMPATYVDAQGKMAHLDVWDSWP 244
DB 364 SSFNTAPNLISN-----NASDIYQNSIYT-----WT 393

QY 245 VQDPVTGVVSNMYGQLVIAAMGIPNSPTGDNHITLYLNKYGDNDFFSHWRN-AGSIFGK 303
DB 394 VQNPW---ACNSM-----NSFNQANNHNGIIPSTT 420

QY 304 ETNV-FQEW-----SGSAIVNDGGTTLQFTSNDTSDYKLNQO--RLATATLNLNV 351
DB 421 SVVDVDFQRCFLGLFDINSQNVICDSQNKNNMLISEHINNVINNERNNRNNGGATLSS 480

QY 352 DDNGVSKSVQDNYQVLFEGGFHYOT--YEQFANGKD-----RENDYDCL--RDPHVQV 401
DB 481 HQNMKPLDSDNNACTGEGDGI SVESGVNAYINGSNLTENKLENRLQIGSNMNDHLIA 540

QY 402 L--ENGDRYLIVFEANTGTEYQSDQDIYNANY-----GGDDAFNIKSSFKLL 447
DB 541 INEDSGGNNI-----NSSINDTNTQTLFNSAPPLTSALIKQIPSSKDTKISSVTSSI 596

QY 448 NKKDRELAGLANGALGILKLTNNQSKPKVEEVYSPVSTLMACDEVXX---KLGDYYL 504
DB 597 NNNN-----GLNTGNSQOECNLSNPSKLS-----ISTSGSNQNNNNNNPSTGSPNKL 646

```

```

QY 505 FSVTRY-SRGSRE-----LTA---KONTI-----VGDNVAMIGYVS---DSLGMKY 544
DB 647 ISMMGYRNEGSIRNNLIGSALSANNKQNTINSGRPGVGVGNLNLSTGNVSYIDAVLHK- 705

QY 545 KPLNNSGVLTASVPANWRRTATYSYAVPVAGHPDQVLITSYMSNKDFASGEGNYATWAP 604
DB 706 NDSNSGGVTTTSS-----STGRSENQASVNSDMSN-----735

QY 605 SELVQINPDPTTIVLARATNQGDWVDDSRNDNMLGVLKEGAANSAAALPGWGKPVDS 664
DB 736 -----ISNDSNNLLSVNTD-----EDSS---TALGAHGEYNYKRALCRHWMRGYCW 781

QY 665 LIN-----RSPGLGLKPHQVPQKIDQPDQPSQNTKNV-----T 700
DB 782 EADCKFAHGEAELRTRDGLRHLPTLSNGSEVANQNGSQSQSQSQSQYSAALQPT 841

QY 701 PNGDKPAGKATPDNTNIDPSAQPSQNTNIDPSAQXGQNTKNVTPGNEKQGN 755
DB 842 VNNVTIPSGNLATSTSTFTSTSGG--SSVSSSIQSKGLKQGS---GNGKNGSN 891

RESULT 13
Q95PH4
ID Q95PH4 PRELIMINARY; PRT; 1318 AA.
AC Q95PH4;
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DE HISTIDINE KINASE DHKM (FRAGMENT).
GN DHKM.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Anjard C., Loomis W.F.;
RT "The histidine kinases of Dictyostelium.";
RL (in) Inouye M., Dutta R. (eds.);
RL HISTIDINE KINASES IN SIGNAL TRANSDUCTION, pp.1-1, Academic press,
RL San Diego (2001).
DR EMBL; AF362374; AAK54093.1; -
KW Kinase.
FT NON_TER 1
SQ SEQUENCE 1318 AA; 144944 MW; B1DA5CA0427066F6 CRC64;

```

```

Query Match 5.0%; Score 207.5; DB 5; Length 1318;
Best Local Similarity 19.5%; Pred. No. 0.024;
Matches 163; Conservative 116; Mismatches 306; Indels 249; Gaps 38;

QY 30 QVESNNYNGVAEVNTERQANGQIGVDGKIISANTSTSGSTNQESSATNTNENAVVNESK 89
DB 229 ELKKSNNDKIIE-NKENQENNNNNNN-----NNNNNNNNNNNNNNNNNNNNNNNN 282

QY 90 NNTNENAVVNEKNKNTNENAVVNEKN-----TNTTENDNSQLKLTNNEQPSAATQANL 145
DB 283 NDSNNNNNCINNDLKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 341

QY 146 KKLNPQAKAVQNAKIDAGSLTDDQINELNKFNSKAEGAKLTKDLE-GIGNAIVKQ 204
DB 342 KKKKKNEFTVFSVIDSGSIDPYSTNLLFQPFSLSS--YNNSTNTDGEFLGLAICKQ 400

QY 205 DPQYAIPTYSNAKEIKNMPATYVDAQGKMAHLDVWDSWPVQ-DPVTGVVSNVMGVLVI 263
DB 401 -----LSNLNGELIKYETENEKSVFELQV-----PMKCDSISSITSSM----- 439

QY 264 AMGIPNSPTG-DNHITLYLNKYGDNDFF--SHW-----RNAGSI-----FG 301
DB 440 -----NSTTNTNTHYPRIMNNQSSKFFANSKWGEGKLITLVDDNPNIGKVIAHLEPG 493

QY 302 TKETNVFQEWSSAIVNDGGTTLQFTSNDTSDYKL-----NDQR- 341

```

```
Db 494 FK-----VFQRTTQSAL-----YFNERNGDFNLILLDPLSLVIDEIKQMQDSSN 542
QY 342 -----LATATL--NLNVDDNGSVTSVDNYQVLFEGDGFHYQTVQEFQANGKDREND 390
Db 543 IIKNPPLVIMCTAKLRKSLNVNDVHYLKPRIKREQL-----TVLSQLLPNTSTINP 593
QY 391 DYCLRDPHVQVLENGDRYLVPEANTGTEDYQSDDOYIYNWANYGGDDAENIKSFKLLNNK 450
Db 594 IYSNQLN-----NSGSSNGGGGGGGGGGGGGSSNNIDF-----NK 634
QY 451 KDRELAGLANGALGILKLTNNQSKPKVEEVYSPVSTLMACDEVXXKLGDKYLYFSVTRV 510
Db 635 TKLGGSNISTG-IGNSGLINSNNIPTVTPSNIIPNLLSCQSL-----677
QY 511 SRGSRELTAKNTIGVNDVAMIGVVSOSLMGKYPKPLNNSGVVLTPASVPANWRTATYSY 570
Db 678 -----LTSLNANIPQLTNDIGITNNN-----INNSLMFTT-----709
QY 571 AVPVAGHPDOVLITSYM-----SNKDFASGEGNYATWAPSLVQINPDPTTVILARATN 624
Db 710 -----PNSTLSNGITGLDNNNSNDTGSIDNNSN-----ISTNDNNNDYFIR--N 753
QY 625 QGDWYMDSSRRDNL-----GVLKEGAANSALPGEWGPVDSLINRSPGLGLKPHQP 679
Db 754 NGIPQNDNTYNNYVLRHQGVLPK-----SLSVP---STPLSYNMLPTNLINAK-----802
QY 680 VQPKIDPQPOQSGGNTKVPNGGDKPAGKATPNT-----NIDPSAQSPSGGNTNIDPSA 735
Db 803 -----RSSLOPLNEN--SVLPTNLTPTILSASPOSLLPMGNDINSILPTQOQSDLOS 854
QY 736 QXSGO-----NTK--NVTPGNEKCKNTDAK-----QLPOTGNSK 768
Db 855 QLELPQLOSTIIRDRGGDILPDSLTLEGQITNLSSNSTISINPLPETNNT 908

RESULT 14
Q97K42 PRELIMINARY; PRT; 2817 AA.
AC Q97K42:
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE UNCHARACTERIZED PROTEIN, RELATED TO ENTEROTOXINS OF OTHER
DE CLOSTRIDIALES.
GN CAC1079.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RA "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum."
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007623; AAK79053.1;
DR InterPro; IPR002479; CW.binding.
DR Pfam; PF01473; CW.binding_1; 119.
KW Complete proteome.
SQ SEQUENCE 2817 AA; 318224 MW; 1851D0D4FFBEE921 CRC64;
```

```
Query Match 4.9%; Score 203.5; DB 16; Length 2817;
Best Local Similarity 21.7%; Pred. No. 0.093;
Matches 143; Conservative 92; Mismatches 211; Indels 213; Gaps 35;
QY 34 NNYNGVAEYNT-----ERQA--NGQIGVDGKIISANTSTSGSTNQESSA--TNN 80
```

```
Db 66 NNLNTVQVNTASSMPNTNPKQATNNSKILVNPFLNQASSPNEGIPPKQASIPYNTV 125
QY 81 ENAVNVESKNTNTNNAVYNE-----NKNTNNT-----ENAVNVENKNTNTN 124
Db 126 DN-----KNTFKNESSINNEAPILPKDTSKTSSAQTSGSNDNNIPSNNTSTNTSKN 179
QY 125 D--NSOLKIT-----NNEQPSAATQANLKLNPQAQAKAVONAKIDAGSLTDD 169
Db 180 ENPSNDITKTTEPANAPITKPTNNQSDSALAKNKAALNNLNDSSQTSKVTSSN-----235
QY 170 QINELNKNFSAEKGAKLTFKLEGIGNAIVKQDPQYAIPTYSNAKEIK-----NMPAT 224
Db 236 --NDAPKVNTTSTDKKASLNNDSDG---WVTKDGKKYV--YVNGVQKGFOSINKSIY 288
QY 225 YTVDD--AOTGKAHLVDVWDSPVQDPVTVGYVSNYMGYQLVIAMGIPNSPTGDNHIYLL 281
Db 289 YFNDGSMQTG-----W-----LKYSNSYFEDASGVML-----TGLQNINGT 326
QY 282 YNKYGNDFSHWRNAGSIFGKTEINVFQEWSSGSAIVNDDGTIOL--PFTSNDTSDYKLNQ 340
Db 327 Y--YGFND-----DGKLLTGLQAIN-----NNYFFNNDGVMTQGTWITCNDKSYFDNNG-374
QY 341 RLATATLNLNVDDNG-----VSIKSDVNYQVLFEG-----DGFHY-----375
Db 375 VMOTGLVHINNKYFGNDGKLLTGLQINNYTYFDSNGVMOTDWITIDGSKYFYSVNG 434
QY 376 --QT--YEQFANGKDRNDYCYLRDPHVQVLENGDRYLVPEAN-----414
Db 435 VMOTGIYISGYYGFA-----NDGKLLTG---LOVINGNSY-YFDTNGIRLSRVWITI 484
QY 415 TGPEDYQSDDOIY--NWANYGGDDAFNIKSSFKLLNNKRELAGLANGALGILKLTNNQ 472
Db 485 DGKDYFENQDGIITDNWYNGKYFYFISG-----VKQTGLQ 521
QY 473 SKPKVEEYSPVSTLMACDEVXXKLGDKYLYFSVTRVSRGSDRELPAKONTIVGDN-VA 531
Db 522 N-----IDGNYFFDSSGIMQTLQKIDGK-TYFVGDNIGIR 556
QY 532 MIGYVSDSLMGKYKPLNNSGVLTASVPANWRTATYSYAVPVAGHPDOVLITSYMSNK 590
Db 557 QIGWITYQ--NKKYY--FNSDGSMDTLKIYSYSPYNYHYQYGFDFNDGKLLTGLQTIK 613

RESULT 15
O86489 PRELIMINARY; PRT; 1166 AA.
ID O86489
AC O86489;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SDR E PROTEIN.
GN SDR E.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NEWMAN;
RX MEDLINE=99098700; PubMed=9884231;
RA Josefsen E., McCrea K., Ni Eidhin D., O'Connell D., Cox J., Hook M.,
RA Foster T.J.;
RA "Three new members of the serine-aspartate repeat protein multigene
RT family of Staphylococcus aureus."
RL Microbiology 144:3387-3395(1998).
DR EMBL; AJ005647; CAA06652.1;
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF00746; Gram_pos_anchor.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN.1.
SQ SEQUENCE 1166 AA; 126348 MW; 750A7B0135287D4A CRC64;
```

Query Match	4.8%;	Score 199.5;	DB 2;	Length 1166;
Best Local Similarity	22.0%;	Pred. No. 0.051;		
Matches 166;	Conservative 99;	Mismatches 312;	Indels 177;	Gaps 39;
Qy	12	TLVASIILMGW-----TAHADQ-----	YESNYNGVAEYNT	44
Db	29	TVGTASIIIVGTLIFGLGNQAKAAEYNTSTNAKODDATTSNDKKEVSETENNSTTENS	88	
Qy	45	ERQANGQIGVGGKIISANSNTSGSTNQESSATNTTNAVYNESKNTNTTNAVNVNENK	104	
Db	89	TNPIKKTNTDSQPEAKKESTSSSTQKQNNVTATTETKPKONIEK-----ENVPSTDKT	143	
Qy	105	TNNTENAVVNEKNNTNTENDNSQLKTNNQPSAAQANLKLNQAKAVQN-----	158	
Db	144	ATEDTSVLEEKKAPNTNNVTTKPTST-----EPS--TSEIQTKPTTQESTNIENSQPOPT	200	
Qy	159	-AKIDAGSLTDDQINELNKNINFSASAKGAKLTPEKDL-EGIGNAIVKODPOYAIYPSNAK	216	
Db	201	PSKVD--NOVTDATNPREPNNVNSKEELKNPELKLVELNRDNTNDHSTKTPVATAPTVP	258	
Qy	217	EIKNMPATYVDAQTGKMAHLVDWD-----SNVPQDPVTGYVSNYNGYOLVIAMGIP	269	
Db	259	KRVNAKRMFAV-AQPAAVASNNVDLIKVTQTKIVGDKGNVAAAHDGKDI-----EYD	312	
Qy	270	NSPFGDNHIIYLLNKYGDN-DFSHRNA--GSIFGTETNVFQWSSGSAIYN---DDGTI	323	
Db	313	TEFTIDNKK-----KKGDTMTINDKNVPSDLTDKNDPIDITDPSGEVIAKGFDTKATK	367	
Qy	324	QLFFTSNDTSD-----YKLNDR-----LATA-----TLNLNVDDNGVS	357	
Db	368	QITVFTDYVDKYEDIKSRLTLYSIDKTVPNETSNLNFTFATAGKTSQNVTVYQDPM	427	
Qy	358	IKSVNTQVLPEGGFHYQTYEQ--FAN--GKDREN-----DDYCLDRPHVVQLE	403	
Db	428	VHGSNQSISQIFTKLDEKQTEQOIYVNPCLKSATNTKVDIAGSOVDY-----GNIKLG	482	
Qy	404	NGDRVLYFEANTGEDYQ--SDQO-----IYNWANGDDDAFIKSFKLLNKK--DR	453	
Db	483	NGS--TIIDQNETKVKVNSDQQLPQSNRIYDFSQYE-----DVTQF---DNKKSFNS	532	
Qy	454	ELAGLANGALIKLTNNQSPKPEEYVSPLV-STLMACDEVXXKLGDKYLFVSWTRVSR	512	
Db	533	NVATLDFGDI-----NSAYIIKVVSVYTPSTDGELDIAQGTSNRTDKYGYNVA----	582	
Qy	513	GSDELTAKNVTIYGDNVA-----MIG-YV-----SDSLMG---KYKPLNNSGVVLTA	556	
Db	583	GYSNFIYTSNDTGGDGTVPKPEEKLYKIGDYWBEDVDKGVQGTDSKEKPNAN--VLVTL	640	
Qy	557	SVP-----ANNRATSYATYVAVPAGHPQVLI-----TSYMSNKDFASEGNYATWAPSEL	607	
Db	641	TYPDGTKSVRTDANGHYERGGGLKDGETYVTKVFTPTPGYLPYTKVNGTDEKKGSSVT	700	
Qy	608	VQINPD-----DFTTVILARATNQGDMYDDSSRN	636	
Db	701	VKINGKDDMSLDTGFKPEKPNILGDYWBEDTNK	734	

Search completed: September 26, 2002, 18:28:01  
Job time: 260 sec